SPATIAL-TEMPORAL MODELS FOR PROCESSES ON THE SPHERE AND THEIR APPLICATION IN CLIMATE PROBLEM

A Dissertation

by

JAEHONG JEONG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Mikyoung Jun
Committee Members,	Suhasini Subba Rao
	Matthias Katzfuss
	Ramalingam Saravanan
Head of Department,	Valen Johnson

August 2015

Major Subject: Statistics

ABSTRACT

There have been noticeable advancements in developing parametric covariance models for spatial and spatial-temporal data in climate science. However, literature on covariance models for processes on the surface of a sphere is still sparse, due to its mathematical difficulties. In this dissertation, we study random fields and spatial-temporal covariance functions on the surface of a sphere. At first, smooth climate variables need smooth covariance functions. We develop a methodology to construct parametric covariance functions using the great circle distance for spatial processes, geared towards smooth processes on the surface of a sphere. We integrate a non-differential process over a small neighborhood on the surface of a sphere, which result in a smoother process. The resulting model is isotropic and positive definite on the surface of a sphere with the great circle distance, with a natural extension for nonstationarity case. Extensive numerical comparisons of our model, with a Matérn covariance model using the great circle distance as well as the chordal distance, are presented.

Next, utilizing the one-to-one mapping between the Euclidean distance and the great circle distance, isotropic and positive definite functions in a Euclidean space can be used as covariance functions on the surface of a sphere. However, this approach may result in physically unrealistic distortion on the sphere especially for large distances. We consider several classes of covariance functions on the surface of a sphere, defined with either the great circle distance or the Euclidean distance, and investigate their impact upon prediction. We demonstrate that covariance functions originally defined in the Euclidean distance may not be adequate for some global data.

Finally, climate variables often vary in both space and time and it has become popular to model multiple processes jointly. We consider the extension of the bivariate Matérn covariance models for spatial-temporal processes on the surface of a sphere. Since data sets have large dimension, a number of challenges arise when performing parameter estimation and prediction. To overcome the computational challenges, we consider the Discrete Fourier Transformation (DFT). We present a method to compute the approximate likelihood efficiently for the case of regularly spaced data of large dimension.

DEDICATION

To my family for their infinite trust and unconditional love.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Mikyoung Jun, for her valuable teaching and excellent mentorship through my studies. I always have been amazed by the patience and understanding she has displayed with me over the past few years. Without her guidance and persistent help, this dissertation would have been impossible.

I wish to thank my committee members, Dr. Suhasini Subba Rao, Dr. Matthias Katzfuss, and Dr. Ramalingam Saravanan, for their willingness to serve on my committee and their helpful comments and suggestions. Additionally, I thank Dr. Michael Longnecker for his generous support and advice through my studies and Dr. James Long for his kind support during the final examination.

I am indebted to my friends, Dr. Junbum Lee, Dr. Won Chang, and Donghyuk Lee. They helped me get through the final stage of this study. Lastly, and most importantly, I extend a sincere thanks to my family for all their love and encouragement.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

1. INTRODUCTION

There have been noticeable advancements in developing parametric covariance models for spatial and spatio-temporal data with various applications to environmental problems. However, literature on covariance models for processes defined on the surface of a sphere with great circle distance as a distance metric is still sparse, due to its mathematical difficulties. This topic is an important issue in the analysis of complex physical processes on the globe such as satellite measurements of climate variables. The theoretical challenges in modeling spatial processes lies in formulating a valid covariance function. For an integer $d \geq 1$, let $\mathcal{S}_r^d = \{\mathbf{x} \in \mathbb{R}^{d+1} : ||\mathbf{x}|| = r\}$ be a (d-dimensional) sphere with radius r, where $||\mathbf{x}||$ is the Euclidean norm of $\mathbf{x} \in \mathbb{R}^{d+1}$, and define the great circle distance on $\mathcal{S}_r^d$ by $\theta(\mathbf{x}, \mathbf{y}) = r \times \arccos(\langle \mathbf{x}, \mathbf{y} \rangle)$, where $\langle \cdot, \cdot \rangle$ denotes the usual inner product on $\mathbb{R}^{d+1}$. One natural and useful construction of valid covariance functions on a sphere is a projection of covariance functions from $\mathbb{R}^{d+1} \times \mathbb{R}^{d+1}$ to $\mathcal{S}_r^d \times \mathcal{S}_r^d$ through the Euclidean distance (equivalently, the chordal distance on $\mathcal{S}_r^d$) expressed in terms of the great circle distance as $2r \sin\{\theta(\mathbf{x}, \mathbf{y})/(2r)\}$ on $\mathcal{S}_r^d$ (Yadrenko, 1983). This approach is attractive because it allows us to use a variety of covariance functions developed in the Euclidean space and preserves the interpretation of important covariance parameters such as scale, range, smoothness, and fractal index. However, the resulting covariance function can significantly distort the local properties of the process for larger values of the great circle distance, i.e., when two points are farther away on a sphere. Gneiting (2013). This motivated the study of random fields and valid covariance functions on $\mathcal{S}_r^d$. In this dissertation, we study random fields and spatial-temporal covariance functions on the surface of a sphere.

It is known that the popular Matérn covariance function, with smoothness parameter greater than 0.5, is not valid for processes on the surface of a sphere with great circle distance. In Section 2, we introduce an approach to produce Matérn-like covariance functions for smooth processes on the surface of a sphere that are valid with great circle distance. The resulting model is isotropic and positive definite on the surface of the sphere with great circle distance, with a natural extension for non-stationarity case. We present extensive numerical comparisons of our model, with a Matérn covariance model using great circle distance as well as chordal distance. We apply our new covariance model class to sea level pressure data, known to be smooth compared to other climate variables, from the CMIP5 climate model outputs.

Utilizing the one-to-one mapping between the Euclidean distance and the great circle distance, isotropic and positive definite functions in a Euclidean space can be used as covariance functions on the surface of a sphere. This approach, however, may result in physically unrealistic distortion on the sphere especially for large distances. In Section 3, we consider several classes of parametric covariance functions on the surface of a sphere, defined with either the great circle distance or the Euclidean distance, and investigate their impact upon spatial prediction. We fit several isotropic covariance models to simulated data as well as real data from NCEP/NCAR reanalysis on the sphere. We demonstrate that covariance functions originally defined with the Euclidean distance may not be adequate for some global data.

In geophysical and environmental sciences, processes such as climate variables often vary in both space and time. Covariance models for space-time data observed at locations on the surface of the Earth should be positive definite crossed with time. Moreover, it has become popular to model multiple processes jointly. Cross-covariance functions are critical for the relationship between distinct variables. In Section 4, we extend multivariate covariance models in Gneiting et al. (2010) for

a space-time process on the surface of a sphere. We focus on the spatially varying variance parameters in the model and present a method to compute the approximate likelihood efficiently for the case of regularly spaced data of large dimension. We also illustrate our covariance models on a bivariate example of temperature and surface pressure.

2. A CLASS OF MATÉRN-LIKE COVARIANCE FUNCTIONS FOR SMOOTH PROCESSES ON A SPHERE*

2.1 Introduction

In geophysical and environmental sciences, data sets often come in a global scale. Covariance models for global data need to be positive definite on the surface of a sphere and their natural distance metric is great circle distance. However, literature on covariance modeling on the surface of a sphere with great circle distance is scarce, since it is mathematically challenging. Schoenberg (1942) presented a classical representation for positive definite functions with great circle distance in terms of Gegenbauer polynomials. Yadrenko (1983) and Yaglom (1987) presented a simple but useful idea of restricting valid isotropic covariance functions in $\mathbb{R}^3$ on the surface of a sphere, which requires the use of chordal distance rather than great circle distance. Huang et al. (2011) examined the validity of several parametric isotropic covariance functions with a great circle distance on the surface of a sphere. Some parametric covariance functions on the surface of a sphere are derived from various constructional approaches (Jun and Stein, 2007, 2008; Huang et al., 2011; Du and Ma, 2012; Du et al., 2013; Gneiting, 2013).

The Matérn covariance function is given as

$$K(t;\sigma^2,\alpha,\nu)=\frac{\pi^{1/2}\sigma^2(\alpha^{2\nu})}{2^{\nu-1}\Gamma(\nu+1/2)}(t/\alpha)^{\nu}\mathcal{K}_{\nu}(t/\alpha),\quad t\geq 0, \tag{2.1}$$

where $\sigma^2>0$, $\alpha>0$, and $\nu>0$ represent marginal variance, spatial range, and smoothness parameters, respectively. Here, $\mathcal{K}_{\nu}$ is the modified Bessel function of the

*Reprinted from *Spatial Statistics*, 11, Jeong, J. and Jun, M., A class of Matérn-like covariance functions for smooth processes on a sphere, 1-18, 2015, with permission from Elsevier.

second kind of order ν. The larger ν is, the smoother the resulting Matérn field is, and smoothness means mean square smoothness (Stein, 1999). In particular, the Matérn field is at least m times mean square differentiable if and only if $\nu > m$, due to the fact that

$$K(t) = \begin{cases} \sum_{j=0}^{m} a_j t^{2j} - \dfrac{\pi\sigma^2}{\Gamma(2\nu+1)\sin(\nu\pi)} t^{2\nu} + O(t^{2m+2}), \text{ as } t \downarrow 0, & \text{if } m < \nu < m+1, \\ \sum_{j=0}^{m} b_j t^{2j} + \dfrac{2(-1)^m \sigma^2}{(2m+2)!} t^{2m+2}\log(t) + O(t^{2m+2}), \text{ as } t \downarrow 0, & \text{if } \nu = m+1, \end{cases} \tag{2.2}$$

for appropriate constants a_j and b_j, $j = 1, \ldots, m$, depending on σ^2, ν, and α (Stein, 1999). The Matérn class is positive definite in $\mathbb{R}^d$ for any $d \in \mathbb{N}$ with the Euclidean distance, i.e., chordal distance on the surface of a sphere (Stein, 1999).

According to Miller and Samko (2001) and Gneiting (2013), the Matérn class using great circle distance is valid if and only if the smoothness parameter satisfies $0 < \nu \leq 0.5$ for 1, 2, and 3 dimensional spheres, and thus it is not mean square differentiable at the origin. This implies that the Matérn class with great circle distance cannot be adapted for smooth processes on the surface of a sphere.

We propose a new class of Matérn-like covariance functions with great circle distance on the surface of a sphere. Resulting covariance function guarantees positive definiteness and is designed to model smooth processes on the surface of a sphere. The new approach integrates a non-differentiable process over a small neighborhood on the surface of a sphere to make it differentiable, and uses finite summation approximation along with practical algorithm for the approximation. Furthermore, by considering local integration vary over space, it can be easily adapted for nonstationary processes.

Hansen et al. (2011) provided a flexible framework with Lévy particles for model-

ing and simulating star-shaped random particles in three dimension. Our covariance functions may be connected to this framework with a certain choice of kernel functions. However, parameter estimation is not discussed in details in the paper and it is not clear how practical the proposed approach is.

Recently, Guinness and Fuentes (2013) proposed the, so called, circular Matérn covariance function in terms of a Fourier series, which is valid on the surface of a sphere with great circle distance. They argued that the circular Matérn covariance function is m times mean square differentiable by defining the notion of mean square differentiable processes on the surface of a sphere. It has a closed-form only when the smoothness parameter is a half-integer. For other smoothness values, it is expressed as an infinite summation. Therefore, in practical application, we may need to truncate the summation at finite terms. To deal with truncation issue for arbitrary smoothness values, they addressed the approximation theorems that provide guarantees about positive definiteness. They also give computationally efficient methods.

The rest of the Section is organized as follows. Section 2.2 presents the new method proposed. It provides an illustration of the method on $\mathcal{S}^1$ and provides details on implementation on $\mathcal{S}^d$ for $d = 1, 2, 3$. Then we present four simulation studies on $\mathcal{S}^2$ in Section 2.3. Section 2.4 gives real application results to a sea level pressure data. Finally, Section 2.5 provides some discussion and future work.

2.2 Method

Our goal is to construct a class of Matérn-like covariance functions for smooth processes on the surface of a sphere, which is valid with great circle distance. To avoid the hassle of verifying the positive definiteness of proposed covariance models, we start from a homogeneous process whose covariance structure is given by a Matérn

class with $0 < \nu \leq 0.5$. Our key idea is to integrate such process over a small neighborhood on the surface of a sphere, which results in a smoother process.

Suppose $Z(\mathbf{s})$, $\mathbf{s} \in \mathcal{S}^d$, $d = 1, 2, 3$ ($\mathcal{S}^d$ is the surface of a sphere with radius R in $\mathbb{R}^{d+1}$), is a homogeneous process on the surface of a sphere, $\text{cov}\{Z(\mathbf{s}), Z(\mathbf{t})\} = K(\theta; \sigma^2, \alpha, \nu)$ with $\theta = \theta(\mathbf{s}, \mathbf{t}) = \arccos(\langle \mathbf{s}, \mathbf{t} \rangle)$, and K given in (2.1). Then let us define a new process on $\mathcal{S}^d$,

$$W_\epsilon(\mathbf{s}) = \frac{1}{\Delta_\epsilon} \int_{\delta_\epsilon(\mathbf{s})} Z(\mathbf{t}) d\mathbf{t}, \tag{2.3}$$

where $\delta_\epsilon(\mathbf{s}) = \{\mathbf{u} \in \mathcal{S}^d : \theta(\mathbf{s}, \mathbf{u}) \leq \epsilon\}$ for some $\epsilon > 0$ and the integral is done on the surface of a sphere. Further, Δ_ϵ denotes the area of $\delta_\epsilon(\cdot)$ on the surface of the sphere (with radius, R). Then, the covariance of W_ϵ is given by,

$$\text{cov}\{W_\epsilon(\mathbf{s}), W_\epsilon(\mathbf{t})\} = \frac{1}{\Delta_\epsilon^2} \int_{\delta_\epsilon(\mathbf{s})} \int_{\delta_\epsilon(\mathbf{t})} K\{\theta(\mathbf{u}, \mathbf{v}); \sigma^2, \alpha, \nu\} d\mathbf{u} d\mathbf{v}, \tag{2.4}$$

where $\sigma^2 > 0$, $\alpha > 0$, $0 < \nu \leq 0.5$, and $\epsilon > 0$ represent marginal variance, spatial range, smoothness, and integration range parameters for Z, respectively. For simplicity, we assume that mean of Z is zero, and thus mean of W_ϵ is zero. If we assume that Z follows Gaussian distribution (which is often done for spatial and spatio-temporal data in practice), then the distribution of W_ϵ is also Gaussian. This is due to the fact that all linear functionals of a Gaussian process, such as derivatives or integrals of Gaussian processes, are Gaussian processes (Parzen, 1999). Note that we normalize the process in (2.3) by Δ_ϵ, so that variance of W_ϵ is the same as that of Z. From now on, we drop ϵ in W_ϵ.

The model proposed in (2.3) may seem similar to a kernel averaged process with a *top-hat filter*, a function shaped similar to a density function of continuous uniform

distribution. Along that line, one may consider ϵ as a bandwidth of such kernel. Note that Heaton et al. (2014) used a specific kernel to model processes on the surface of a sphere. However, the motivation of integrating a process in (2.3) is quite different from kernel averaging. The main purpose for the integration in (2.3) is to achieve a smoother process. One may fix ϵ at a very small value to achieve a smooth W. In fact, moderate or large values of ϵ may result in undesirably strong spatial dependence of W at a small scale, which may cause numerical problems in dealing with covariance matrix of W. Adding a kernel function in (2.3) results in a complex form of covariance function in (2.4), which is not desirable in our case. To achieve an even smoother process, we may simply keep repeating the integrals in (2.3). See (2.10) in Section 2.3.2 for further details.

2.2.1 Illustration on $\mathcal{S}^1$

In $\mathcal{S}^1$, the analytic calculation of the double integrals in (2.4) is relatively simple. Consider a process $W(s) = \int_{s-\epsilon}^{s+\epsilon} Z(u)\, du/(2\epsilon)$ on $\mathcal{S}^1$. In particular, when K is an exponential function with range α, we can easily obtain an explicit expression for the covariance function of W. Some tedious calculation gives us ($\theta = \theta(s,t) = |t-s|$):

$$
\begin{aligned}
&(2\epsilon)^2 \mathrm{cov}\{W(s), W(t)\} \\
&= \begin{cases}
\alpha(4\epsilon - 2\theta) + \alpha^2(-2e^{-\theta/\alpha} + e^{\frac{\theta-2\epsilon}{\alpha}} + e^{\frac{-\theta-2\epsilon}{\alpha}}), & \text{if } \ \theta \le 2\epsilon, \\
\alpha^2 e^{-\theta/\alpha}(e^{2\epsilon/\alpha} + e^{-2\epsilon/\alpha} - 2), & \text{if } \ 2\epsilon < \theta < \pi - \epsilon, \\
\alpha^2\{-2e^{\frac{\epsilon-\pi}{\alpha}} + 2e^{\frac{-\epsilon-\pi}{\alpha}} + (e^{\frac{-\theta}{\alpha}} + e^{\frac{\theta-2\pi}{\alpha}})(e^{2\epsilon/\alpha} - 1)\}, & \text{if } \ \pi - \epsilon \le \theta \le \pi.
\end{cases}
\end{aligned} \tag{2.5}
$$

Taylor expansion of the function in (2.5) near the origin gives

$$
\mathrm{cov}\{W(s), W(t)\} = b_0 + b_1\theta^2 + b_2\theta^3 + O(\theta^4), \ \text{as } \theta \downarrow 0, \tag{2.6}
$$

for appropriate constants b_0, b_1, and b_2, depending on ϵ and α. Thus, the covariance function of W has 2 derivatives at the origin. It is clear from (2.6) that the main factor that determines the smoothness of W is not ϵ, but the fact that W is defined as an integral of Z.

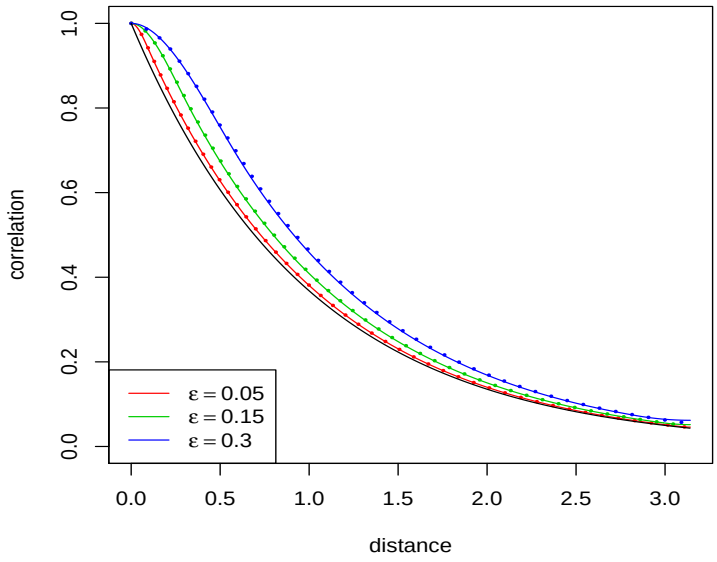

Figure 2.1: The exponential correlation function with marginal variance 1 and spatial range parameter $\alpha = \pi$ (black line), integrated exponential correlation functions given in (2.5) with $\epsilon = 0.05, 0.15$ and 0.3 (straight line) and its approximation as in (2.8) (dotted line).

Figure 2.1 presents the original correlation function of Z, $\rho(\theta) = e^{-\theta/\alpha}$, and the integrated correlation functions of W for $\epsilon = 0.05, 0.15$, and 0.3. Variabilities of the integrated correlation functions decrease slower near the origin than the original correlation function, and the integrated correlation functions look smoother at the origin compared to the original correlation function of Z.

As a sanity check, we simulated Gaussian random fields on $\mathcal{S}^1$ using the proposed covariance model in (2.5), and compared fitted results using the true covariance model (we call it MI*. We use $*$, since by using exponential model, ν for Z is fixed at 0.5. See Section 2.3.1 for details on the notation) as well as Matérn covariance functions with chordal distance (MC) and great circle distance (MG). We considered the spatial range $\alpha = \pi/2$ and π and $\epsilon = 0.03$. We randomly selected 100 locations on a unit circle, and among them, we used 90 randomly selected data points for parameter estimation and the remaining 10 data points for prediction. We repeated this procedure 100 times.

Table 2.1: (Simulation example on $\mathcal{S}^1$) Sample means and standard deviations of parameter estimates and maximum loglikelihood values for each model (100 cases). True covariance function is the integrated exponential covariance given in (2.5) with great circle distance. We set $\alpha = \pi/2, \pi$ and $\epsilon = 0.03$.

α	Model	$\hat{\epsilon}$	$\hat{\alpha}$	$\hat{\nu}$	Max.loglik
$\pi/2$	MC	–	0.330(0.118)	1.169(0.167)	143.680(12.478)
	MG	–	2.906(0.554)	0.500(0.000)	121.532(9.691)
	MI*	0.030(0.005)	1.765(0.441)	-	145.898(22.363)
π	MC	–	0.515(0.158)	1.100(0.127)	169.096(11.730)
	MG	–	5.236(0.812)	0.500(0.000)	150.088(8.924)
	MI*	0.029(0.006)	3.086(0.724)	-	178.642(11.616)

Tables 2.1 and 2.2 show sample means and standard deviations for the model parameter estimates through maximum likelihood estimation, RMSPE, MAE, and CRPS under each model. We observed that the range parameter α is underestimated for MC and overestimated for MG. While all models have comparable performance in prediction, the true model, MI, outperforms MG, in terms of maximum loglikelihood values. This result is not surprising, since MG is not suitable for smooth processes

due to the restriction on smoothness parameter ($\nu \leq 0.5$). On the other hand, MC performs reasonably well despite the distortion at large distances. We found that there is negative association between the estimates of ϵ and α. This is due to the fact that larger ϵ values tend to result in larger effective range. Thus, when the estimate of ϵ is larger than the truth, we tend to get smaller estimates of α to compensate for it. We tested cases with smaller true α values and there was not any noticeable change, in terms of how the three models compare.

Table 2.2: (Simulation example on $\mathcal{S}^1$) Sample means and standard deviations prediction measures for each model (100 cases).

α	Model	RMSPE	MAE	CRPS
$\pi/2$	MC	0.109(0.055)	0.072(0.034)	0.051(0.023)
	MG	0.116(0.053)	0.079(0.033)	0.059(0.024)
	MI*	0.105(0.053)	0.069(0.030)	0.049(0.022)
π	MC	0.075(0.042)	0.050(0.026)	0.036(0.018)
	MG	0.076(0.036)	0.052(0.022)	0.039(0.016)
	MI*	0.068(0.035)	0.045(0.020)	0.032(0.015)

When the smoothness of Z is $0 < \nu \leq 0.5$, due to (2.2), we achieve the following result. As $\theta = |t - s| \downarrow 0$,

$$\text{cov}\{W(s), W(t)\} = a_0 + a_1\theta^2 + c_0\theta^{2(\nu+1)} + O(\theta^4), \tag{2.7}$$

for $c_0 = \frac{-\pi\sigma^2}{2\epsilon^2\Gamma\{2(\nu+1)+1\}\sin\{(\nu+1)\pi\}}$, and appropriate constants a_0, a_1, depending on σ^2, ν, α, and ϵ. Theorem 2 of Stein (1999) and (2.7) imply that the covariance function of W has 2 derivatives at the origin.

2.2.2 Practical approximation of integral for $\mathcal{S}^d$

In general, analytically evaluating the integral in (2.4) on the surface of the sphere is tricky, except for those simple situations presented in Section 2.2.1. Instead, we use a finite approximation of the integrals as in the following ($\mathbf{s}, \mathbf{t} \in \mathcal{S}^d$, $d = 1, 2, 3$):

$$\text{cov}\{W(\mathbf{s}), W(\mathbf{t})\} \approx \frac{1}{N^2} \sum_{\{\mathbf{u}_i \in \tilde{\delta}_\epsilon(\mathbf{s})\}} \sum_{\{\mathbf{v}_j \in \tilde{\delta}_\epsilon(\mathbf{t})\}} K\{\theta(\mathbf{u}_i, \mathbf{v}_j)\}, \tag{2.8}$$

where $K(\cdot)$ is a Matérn covariance function in (2.1) with $0 < \nu \leq 0.5$ and $i, j = 1, \ldots, N$. Here, $\tilde{\delta}_\epsilon(\mathbf{s})$ is a set of a finite number of locations defined as in the following. First, consider n_c number of circles with a center $\mathbf{s}$ and radii, $\epsilon, \epsilon/2, \epsilon/4, \ldots, \epsilon/2^{n_c-1}$. Then on each circle, we consider n_p as many equally spaced (in terms of angles)

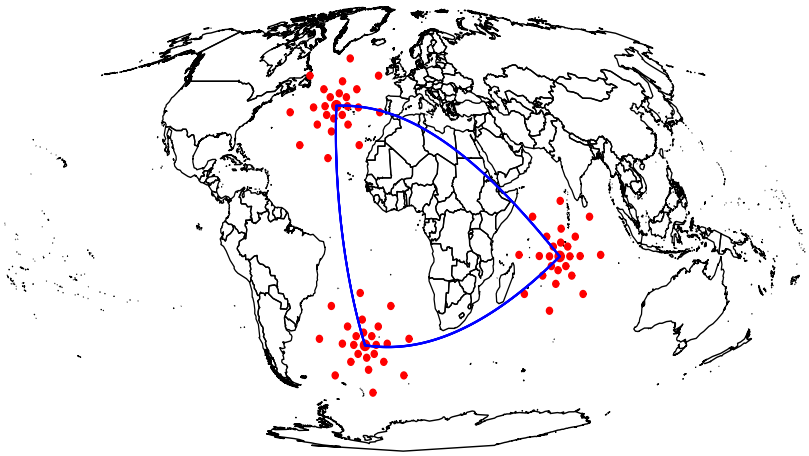

Figure 2.2: The locations used for the approximation, for the centers $(-35°, 40°)$, $(70°, -10°)$, and $(-20°, -40°)$ with great circle distance 2000 (km) as a radius. The number of circles and the number of points on each circle are 3 and 8, respectively.

points. Thus points on each circle are $2\pi/n_p$ radians apart. Then the set $\tilde{\delta}_\epsilon(\mathbf{s})$ consists of $\mathbf{s}$ and n_p points on the circles; $\tilde{\delta}_\epsilon(\mathbf{s})$ contains $n_c \times n_p + 1$ points. Figure 2.2 gives an example of the choice of the finite number of locations used in the approximation in (2.8), and there are 8 points on each circle. Algorithm shown in Figure 2.3 returns a point that is on a circle with specified radius (from the center) and on a circle, at a specified angle apart (clockwise) from the vertical line. When we compute the covariance between two locations on the surface of a sphere, we use a finite approximation of the integrals in (2.8) by averaging covariances among all possible pairs of $n_c \times n_p + 1$ locations. Note that for any n_c and n_p, (2.8) gives a positive definite covariance function, as it is simply a summation of a finite number of positive definite functions.

Instead of what is proposed in Algorithm shown in Figure 2.3, one may choose

---

procedure NEW_LOC(l_1, L_1, θ, ϕ)
  convert degrees to radians $(l_1, L_1, \phi) = (l_1, L_1, \phi) \times \pi/180$
  convert great circle distance to radian $d = \theta/R$
  $L_2 = \arcsin\{\sin(L_1) \times \cos(d) + \cos(L_1) \times \sin(d) \times \cos(\phi)\}$
  $a = \sin(d) \times \sin(\phi)$
  $b = \cos(L_1) \times \cos(d) - \sin(L_1) \times \sin(d) \times \cos(\phi)$
  if (b=0) **then**
    $l_2 = l_1$
  else
    $l_2 = l_1 + \arctan 2(a, b)$
  end if
  convert radians to degrees $(l_2, L_2) = (l_2, L_2) \times 180/\pi$
  return l_2 and L_2
end procedure

Figure 2.3: (Algorithm) Great circle waypoint. This algorithm returns a point that is a specified great circle distance apart from the center and at a specified angle apart from the center (clockwise) (l : longitude, L : latitude, θ : great circle distance, ϕ : angle), and R : radius of the sphere.

$\mathbf{u}_i$'s and $\mathbf{v}_j$'s randomly on each circle. However, for small n_c and n_p, randomly selecting locations of $\mathbf{u}_i$'s and $\mathbf{v}_j$'s may not be effective. When n_p is small, most points may end up close to each other, and thus the finite summation in (2.8) may approximate only a small part of the domain of the integral. Our code to calculate the approximation in (2.8), originally written in C, is available for usage in the software R, and is available at http://www.stat.tamu.edu/~mjun/jeong_jun.html.

2.2.3 Choice of ϵ, n_c, and n_p

The proposed approach requires choices of ϵ, the number of circles n_c, and the number of points on each circle n_p. When $\epsilon = 0$, the integrated covariance function reduces to the covariance function of Z, which is suitable for rough processes. To choose ϵ adaptively, we estimate ϵ along with the parameters of the Matérn covariance function, through the maximum likelihood method. As shown in the example in Section 2.2.1, the estimation of ϵ along with other covariance parameters works well with small n_c and n_p.

The choice of n_c and n_p involves a trade-off between the computational cost and accuracy of approximation. According to our experience, small numbers, such as $n_c = 1$ or 2 and $n_p = 8$ or so, give a reasonable approximation in terms of model fit and spatial prediction. For example, we tested the approximation for the model in Section 2.2.1. For each location $s \in \mathcal{S}^1$, we used 6 points, $s \pm \epsilon$, $s \pm \epsilon/2$, and $s \pm \epsilon/4$, to approximate $\text{cov}\{W(s), W(t)\}$. The dotted lines of Figure 2.1 represent approximated correlation functions with $\epsilon = 0.05, 0.15$, and 0.3. True lines from (2.5) and their approximation through finite summations are nearly identical.

2.2.4 Extension for nonstationarity

In real applications, the assumption of isotropy may be too limited. Data in geophysical applications often exhibit strong dependence of covariance structure on

latitude (Jun and Stein, 2008), and the spatial range or smoothness parameters may vary across space (Kleiber and Nychka, 2012; Jun, 2014). A straightforward extension of our model for nonstationarity may be letting ϵ vary over space. For example, we may allow ϵ to be different over the land and the sea for some climate variables, or to vary across latitude. Jun (2014) found that surface temperature and precipitation variables exhibit strong dependence of their covariance structure on land/sea factor and latitude. See Sections 2.3.3 and 2.4.2 for examples where ϵ is allowed to be different over the land and the sea.

2.3 Simulation studies

We present three simulation studies for the case on $\mathcal{S}^2$, and compare the proposed covariance model to Matérn covariance models with chordal distance as well as great circle distance. We also present comparison among the models using great circle distance in terms of differentiability at the origin. In the first example (Section 2.3.1), the truth is generated from a Matérn model with chordal distance. In the second example (Section 2.3.2), the truth is generated from compactly supported covariance functions with great circle distance. In the third example (Section 2.3.3), a non-stationary Matérn covariance model using chordal distance is used as the truth. In the final example (Section 2.3.4), non-differentiable model, as well as differentiable models (once and twice differentiable) using great circle distance are used as the truth.

2.3.1 Simulation I

We consider a Gaussian random field on the surface of the Earth (with radius $R = 6371$ (km)) with mean 0 and a Matérn covariance function in (2.1) with chordal distance. Note that chordal distance between the two locations, (L_1,l_1) and (L_2,l_2),

on $\mathcal{S}^2$ (L and l denote latitude and longitude, respectively) is given by

$$\mathrm{ch}(L_1, L_2, l_1 - l_2) = 2R\left\{\sin^2\left(\frac{L_1 - L_2}{2}\right) + \cos L_1 \cos L_2 \sin^2\left(\frac{l_1 - l_2}{2}\right)\right\}^{1/2}.$$

Great circle distance between the two locations then is given by $\theta = \mathrm{gc}(L_1, L_2, l_1 - l_2) = 2R \cdot \arcsin\{\mathrm{ch}(L_1, L_2, l_1 - l_2)/(2R)\}$. We set $\sigma^2 = 1$, $\alpha = 1000$ (km), and $\nu = 0.5, 0.75, 1.5, 2.5$. We randomly selected 500 locations over the surface of the Earth. Among 500 data points, we used 450 randomly chosen data points as training data for parameter estimation, and remaining 50 data points as test data for prediction. If we randomly select locations on the Earth based on angles, we may end up with clusters near the poles, since two points with the same angular distance apart near the poles are much closer to each other, compared to two points near the equator. To obtain spatially more regular sample locations over the globe, we sampled locations with percentages depending on the latitude, roughly proportional to the area of each latitude band. That is, we use the percentages, (25.88, 24.12, 20.71, 15.89, 9.99, 3.41), for latitudes, ($\pm$[0,15], $\pm$(15,30], $\pm$(30,45], $\pm$(45,60], $\pm$(60,75], $\pm$(75,90]) (degrees), respectively.

We used the maximum likelihood estimation method for each model with the optim function of the software R. We compared fitted results from Matérn model with chordal distance (MC) to those from Matérn model with great circle distance (MG), as well as those from the proposed model (MI). For MI, we use the notation MIx.y to note that $n_c = x$ and $n_p = y$ used in the approximation of the integrals. Often for the proposed model MI, the estimated ν values are close to 0.5 (particularly for smooth processes). Thus, we fix $\nu = 0.5$ to save computational time. For such cases, we denote by MIx.y*. For MC, $\nu > 0$ is enough to guarantee positive definiteness, and for MG and MI, we need $0 < \nu \leq 0.5$. We compared the models in terms of

prediction, and used RMSPE and two popular scoring rules (Gneiting and Raftery, 2007), the mean absolute error (MAE) and the continuous ranked probability score (CRPS), as the criteria for comparison. We repeat each simulation 100 times.

Table 2.3: (Simulation example I) Sample means and standard deviations of parameter estimates and maximum loglikelihood values for each model (100 cases). True covariance function is the Matérn model with chordal distance. We set $\sigma^2 = 1$, $\alpha = 1000$, and $\nu = 0.5, 0.75, 1.5, 2.5$.

ν	Model	$\hat{\epsilon}$	$\hat{\alpha}$	$\hat{\nu}$	Max.loglik
0.5	MC	–	1035.220(256.59)	0.512(0.09)	-121.483(14.79)
	MG	–	1106.760(210.19)	0.471(0.05)	-121.697(14.81)
	MI1.8	19.827(28.34)	1117.439(274.74)	0.445(0.09)	-121.338(14.74)
	MI2.8	23.285(33.76)	1127.281(303.81)	0.445(0.09)	-121.309(14.72)
0.75	MC	–	1015.962(194.79)	0.763(0.10)	-45.214(16.41)
	MG	–	1643.492(144.74)	0.500(0.00)	-50.310(16.74)
	MI1.8	95.919(53.32)	1393.051(214.90)	0.469(0.06)	-45.524(16.49)
	MI2.8	103.614(57.70)	1410.420(222.15)	0.476(0.06)	-45.744(16.55)
1.5	MC	–	1015.501(117.40)	1.500(0.14)	203.799(17.50)
	MG	–	4387.649(340.74)	0.500(0.00)	136.526(18.42)
	MI1.8	382.907(46.42)	2173.389(257.68)	0.500(0.00)	197.553(17.43)
	MI2.8	464.108(64.94)	2077.739(296.72)	0.500(0.00)	199.878(17.66)
	MI1.8*	382.947(46.66)	2173.238(258.16)	–	197.552(17.43)
2.5	MC	–	1003.867(77.65)	2.513(0.17)	550.428(19.45)
	MG	–	9804.716(896.55)	0.500(0.00)	318.342(23.09)
	MI1.8	573.875(25.74)	5023.835(456.25)	0.500(0.00)	486.069(18.75)
	MI2.8	770.629(35.40)	4185.538(406.43)	0.500(0.00)	512.897(18.69)
	MI1.8*	573.799(25.91)	5023.620(456.58)	-	486.071(18.75)

Table 2.3 displays sample means and standard deviations for the model parameter estimates under each model. When $\nu = 0.5$, the estimates of range parameter α, and the smoothness parameter ν are close to the true Matérn covariance model using chordal distance. Maximum loglikelihood values and prediction measures for each model, are similar, as expected. When $\nu \geq 1.5$, we observe that the range parameter

α is overestimated and the smoothness parameter ν is estimated close to 0.5 for MG. From Table 2.4, we observe that MG has the largest RMSPE, MAE, and CRPS values for $\nu \geq 1.5$, and those values for MC and MI models are comparable. When $\nu = 0.75$, we observe similar results from the cases of $\nu \geq 1.5$. The proposed model has larger maximum loglikelihood value than MG and it has similar performance in prediction to MC. We suspect similar results for $n - 0.5 < \nu < n, n \in \mathbb{N}$. These results suggests that the proposed approach performs significantly better than MG for smooth data, and it has similar performance in prediction to MC.

As we mentioned before, the proposed approach involves a trade-off between the

Table 2.4: (Simulation example I) Sample means and standard deviations of prediction measures for each model (100 cases). True covariance function is the Matérn model with chordal distance. We set $\sigma^2 = 1$, $\alpha = 1000$, and $\nu = 0.5, 0.75, 1.5, 2.5$.

ν	Model	RMSPE	MAE	CRPS
0.5	MC	0.714(0.080)	0.562(0.064)	0.398(0.042)
	MG	0.713(0.081)	0.562(0.064)	0.398(0.043)
	MI1.8	0.714(0.081)	0.563(0.065)	0.399(0.043)
	MI2.8	0.714(0.080)	0.563(0.064)	0.399(0.043)
0.75	MC	0.575(0.074)	0.446(0.062)	0.317(0.041)
	MG	0.579(0.075)	0.449(0.061)	0.319(0.041)
	MI1.8	0.577(0.075)	0.448(0.062)	0.318(0.040)
	MI2.8	0.577(0.074)	0.448(0.061)	0.319(0.040)
1.5	MC	0.288(0.040)	0.212(0.027)	0.149(0.018)
	MG	0.325(0.044)	0.242(0.030)	0.174(0.020)
	MI1.8	0.297(0.041)	0.223(0.028)	0.159(0.019)
	MI2.8	0.299(0.041)	0.222(0.028)	0.162(0.017)
	MI1.8*	0.297(0.041)	0.223(0.028)	0.159(0.019)
2.5	MC	0.114(0.022)	0.078(0.013)	0.055(0.009)
	MG	0.177(0.029)	0.126(0.018)	0.097(0.011)
	MI1.8	0.137(0.021)	0.104(0.015)	0.076(0.010)
	MI2.8	0.131(0.022)	0.098(0.014)	0.075(0.008)
	MI1.8*	0.137(0.021)	0.104(0.015)	0.076(0.010)

computational cost and accuracy of inference. According to a preliminary study, the performance of the proposed model improves significantly by increasing n_c and n_p for smooth data. Note that MI1.8 and MI1.8* work similarly in terms of maximum likelihood estimation and prediction.

2.3.2 Simulation II

We now use the following compactly supported covariance functions using great circle distance, $\theta \in [0, \pi R]$, which are defined by

$$
\begin{aligned}
K_1(\theta; \alpha, \tau) &= (1 + \tau\tfrac{\theta}{\alpha})(1 - \tfrac{\theta}{\alpha})_+^{\tau}, \quad \alpha \leq \pi R;\ \tau \geq 4, \\
K_2(\theta; \alpha, \tau) &= \{1 + \tau\tfrac{\theta}{\alpha} + \tfrac{\tau^2-1}{3}(\tfrac{\theta}{\alpha})^2\}(1 - \tfrac{\theta}{\alpha})_+^{\tau}, \quad \alpha \leq \pi R;\ \tau \geq 6,
\end{aligned} \tag{2.9}
$$

where K_1 and K_2 are C^2-Wendland and C^4-Wendland covariance functions (Gneiting, 2013), respectively, and for any function g, $g(\theta)_+ = \max\{g(\theta), 0\}$. Here, α is a range parameter, and τ is a shape parameter. Both covariance functions are known to be valid on a sphere with great circle distance (Gneiting, 2013). We let $\alpha = \pi R \approx 20015.09$ (km), $\tau = 14$ for K_1, and $\tau = 16$ for K_2. We randomly selected 100 locations over the surface of the Earth. Among 100 data points, we used 90 randomly chosen data points for parameter estimation and 10 points for prediction.

For the C^2-Wendland case, Tables 2.5 and 2.6 show that the proposed model performs as well as MC, in terms of maximum loglikelihood and prediction. Note that C^2-Wendland function has two derivatives at the origin. We find that both MC and MI perform equally well, and outperform MG. For the C^4-Wendland case, there is a significant gap in maximum loglikelihood value between MC and MI. This may be due to the fact that C^4-Wendland has four continuous derivatives at the origin. To verify this, we now use double integration in defining the process (call this MII),

that is, consider a new process V such that

$$V_{\epsilon_2}(\mathbf{s}) = \frac{1}{\Delta_{\epsilon_2}} \int_{\delta_{\epsilon_2}(\mathbf{s})} W_{\epsilon_1}(\mathbf{t}) d\mathbf{t}, \tag{2.10}$$

with W_{ϵ_1} as defined in (2.3). It is clear that MII outperforms MI, and fitted results

Table 2.5: (Simulation example II) Sample means and standard deviations of parameter estimates and maximum loglikelihood values for each model (100 cases). True covariance functions are the C^2-Wendland function with great circle distance ($\alpha = 20015.09$ and $\tau = 14$), and the C^4-Wendland function with great circle distance ($\alpha = 20015.09$ and $\tau = 16$).

	Model	$\hat{\epsilon}$ or $\hat{\epsilon}_1/\hat{\epsilon}_2$	$\hat{\alpha}$	$\hat{\nu}$	Max.loglik
C^2	MC	–	1343.405(359.06)	1.641(0.31)	73.103(7.24)
	MG	–	8074.816(1564.76)	0.500(0.00)	51.208(8.93)
	MI3.6	785.728(130.84)	2803.724(876.84)	0.498(0.01)	71.195(7.15)
C^4	MC	–	1100.667(194.21)	2.824(0.45)	137.071(8.32)
	MG	–	14315.879(3752.58)	0.500(0.00)	75.501(11.59)
	MI3.6	1127.589(87.97)	5496.539(1378.87)	0.498(0.00)	119.530(9.01)
	MII3.6	736.812(155.27) 1042.041(168.57)	2952.331(944.47)	0.500(0.00)	134.284(8.24)

Table 2.6: (Simulation example II) Sample means and standard deviations of prediction measures for each model (100 cases). True covariance functions are the C^2-Wendland function with great circle distance ($\alpha = 20015.09$ and $\tau = 14$), and the C^4-Wendland function with great circle distance ($\alpha = 20015.09$ and $\tau = 16$).

	Model	RMSPE	MAE	CRPS
C^2	MC	0.190(0.080)	0.141(0.053)	0.100(0.039)
	MG	0.234(0.091)	0.173(0.063)	0.128(0.043)
	MI3.6	0.192(0.080)	0.142(0.050)	0.100(0.036)
C^4	MC	0.078(0.045)	0.052(0.023)	0.037(0.016)
	MG	0.130(0.071)	0.093(0.042)	0.075(0.028)
	MI3.6	0.087(0.052)	0.059(0.024)	0.043(0.018)
	MII3.6	0.081(0.049)	0.054(0.024)	0.038(0.017)

from MII are more comparable to those from MC, compared to MI. It is evident that doing double integration of the Matérn field is helpful for smooth data with C^4-Wendland covariance function.

2.3.3 Simulation III

We used the nonstationary Matérn covariance model that allows for spatially varying local dependence through ν (Paciorek and Schervish, 2006; Kleiber and Nychka, 2012). We let ν to be different over the sea and the land. We used chordal distance as the distance metric and set $\sigma^2 = 1, \alpha = 1000$ (km), $\nu_{sea} = 1.5$, and $\nu_{land} = 0.5$. We randomly selected 1000 locations over the surface of the Earth similarly to Section 2.3.1, and they were used for parameter estimation. For prediction, we also randomly selected 100 locations over four distinct regions (25 locations for each region). One land region near the U.S. states of Colorado is situated between latitudes 38°N and 39°N, and longitudes 105°W and 106°W. The other land region near Germany is situated between latitudes 50°N and 51°N, and longitudes 10°E and 11°E. Two sea regions are located in South Pacific Ocean (latitudes 13°S and 14°S, and longitudes 170°W and 171°W near Samoa island) and Indian Ocean (latitudes 9°S and 10°S, and longitudes 70°E and 71°E) (see Figure 2.4).

We fitted a nonstationary version of the proposed model (call it, NMI), as discussed in Section 2.2.4. We let ϵ be different over the sea and the land. Table 2.7 shows that the NMI2.8 outperforms all stationary covariance models (MC, MG, and MI2.8) in terms of maximum loglikelihood values significantly. Stationary covariance models lead to inadequate model fit. They are unable to capture the local variability because of the global correlation structure. However, by varying ϵ of NMI, we are able to capture this behavior well.

For prediction, all covariance models have similar performance regarding RMSPE

and MAE, but NMI2.8 has the smallest mean CRPS (see Table 2.8). Moreover, NMI2.8 outperforms in terms of prediction measures over the sea. On the land, all

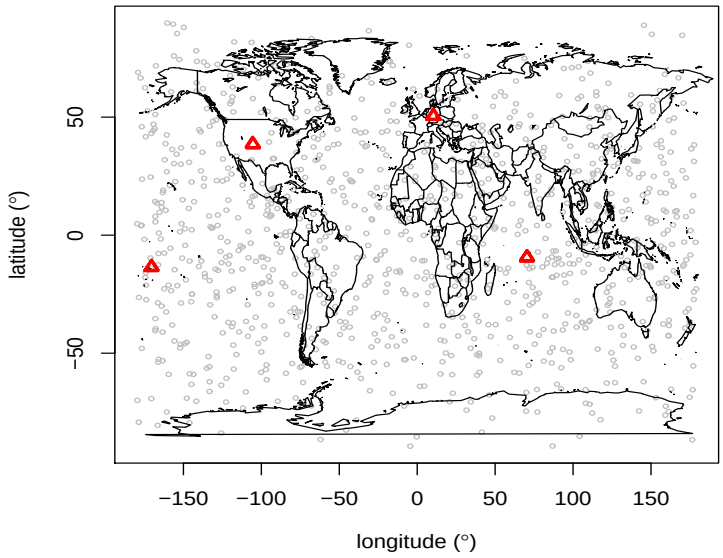

Figure 2.4: (Simulation example 3) One realization of sampling locations, estimation sites (∘) and prediction regions (△) for Section 2.3.3. Each prediction region consists of 25 locations.

Table 2.7: (Simulation example III) Sample means and standard deviations of parameter estimates and maximum loglikelihood values for each model (100 cases). True covariance function is the nonstationary Matérn model with chordal distance. We set $\sigma^2 = 1$, $\alpha = 1000$, $\nu_{sea} = 1.5$, and $\nu_{land} = 0.5$. For NMI2.8, $\hat{\epsilon}_{sea/land}$, and for prediction measures, $\text{RMSPE}_{sea/land}$, $\text{MAE}_{sea/land}$, and $\text{CRPS}_{sea/land}$ are displayed.

Model	$\hat{\epsilon}$	$\hat{\sigma}^2$	$\hat{\alpha}$	$\hat{\nu}$	Max.loglik
MC	-	1.22(0.18)	1936.72(601.03)	0.62(0.08)	131.48(39.50)
MG	-	1.22(0.27)	3111.12(764.36)	0.50(0.01)	123.54(36.58)
MI2.8	37.03(30.34)	1.18(0.19)	2644.21(512.35)	0.50(0.00)	131.51(38.45)
NMI2.8	502.33(31.04) 0.31(0.63)	1.26(0.16)	1434.01(178.69)	0.50(0.02)	419.50(32.90)

covariance models show similar prediction performances. This may imply that NMI captures the local variability over the sea and the land well by means of spatially varying ϵ.

2.3.4 Simulation IV

We used three mean zero Gaussian random fields on the surface of the Earth with non-differentiable model, once-, and twice-differentiable models using great circle distance (i.e. MG, MI, and MII, in our notation). We set $\sigma^2 = 1$, $\alpha = 2000$ (km), $\nu = 0.5$ for all models, $\epsilon = 100$ (km) for MI, and $\epsilon_1 = 200$ and $\epsilon_2 = 100$ (km) for MII. For each data generated, we estimated and performed prediction with non-differentiable, once-, and twice-differentiable models. We randomly selected 100 locations over the surface of the Earth. We used 90 randomly chosen data points for parameter estimation and 10 points for prediction. We repeated this process 100 times.

Tables 2.9 and 2.10 display sample means and standard deviations of the max-

Table 2.8: (Simulation example III) Sample means and standard deviations of prediction measures for each model (100 cases). True covariance function is the nonstationary Matérn model with chordal distance. We set $\sigma^2 = 1$, $\alpha = 1000$, $\nu_{sea} = 1.5$, and $\nu_{land} = 0.5$. For NMI2.8, $\hat{\epsilon}_{sea/land}$, and for prediction measures, $\text{RMSPE}_{sea/land}$, $\text{MAE}_{sea/land}$, and $\text{CRPS}_{sea/land}$ are displayed.

Model	RMSPE	MAE	CRPS
MC	0.191(0.136)	0.162(0.113)	0.140(0.057)
	0.635(0.276)	0.544(0.258)	0.401(0.207)
MG	0.202(0.137)	0.172(0.114)	0.146(0.059)
	0.637(0.279)	0.546(0.260)	0.400(0.208)
MI2.8	0.195(0.136)	0.165(0.113)	0.142(0.059)
	0.636(0.276)	0.545(0.259)	0.400(0.205)
NMI2.8	0.166(0.123)	0.143(0.101)	0.103(0.069)
	0.647(0.276)	0.560(0.260)	0.392(0.180)

Table 2.9: (Simulation example IV) Sample means and standard deviations of parameter estimates and maximum loglikelihood values for each model (100 cases). True covariance functions are MG($\sigma^2 = 1$, $\alpha = 2000$, $\nu = 0.5$), MI1.8($\sigma^2 = 1$, $\alpha = 2000$, $\nu = 0.5, \epsilon = 100$), and MII1.4($\sigma^2 = 1$, $\alpha = 2000$, $\nu = 0.5, \epsilon_1 = 200$, $\epsilon_2 = 100$). For MII1.4, $\hat{\epsilon}_1/\hat{\epsilon}_2$ is displayed.

	Model	$\hat{\epsilon}$	$\hat{\alpha}$	$\hat{\nu}$	Max.loglik
MG	MG	·	2635.740(1887.05)	0.456(0.08)	-18.58(6.77)
	MI1.8	2.38(23.68)	2033.525(435.98)	0.500(0.00)	-18.76(6.72)
	MII1.4	0.27(2.60) 2.37(23.60)	2034.168(436.11)	0.499(0.01)	-18.76(6.72)
MI1.8	MG	·	2631.445(960.92)	0.484(0.05)	-14.05(7.08)
	MI1.8	161.60(147.36)	2187.339(1065.17)	0.464(0.07)	-12.87(7.20)
	MII1.4	207.31(268.33) 148.02(214.01)	3012.419(4834.41)	0.425(0.15)	-13.22(7.28)
MII1.4	MG	·	2932.159(1015.88)	0.495(0.03)	-9.49(7.14)
	MI1.8	258.45(170.96)	2099.711(1091.51)	0.497(0.03)	-7.16(7.16)
	MII1.4	230.79(163.26) 154.54(124.80)	2042.154(687.64)	0.480(0.05)	-6.86(7.18)

imum loglikelihood estimates and prediction measures for each model. As we expected, for the data generated from non-differentiable model, all models fitted have similar performance in terms of maximum loglikelihood values and prediction. For the data generated using the model that is only once differentiable, non-differentiable model performed worse than differentiable models in terms of estimation and prediction. For the data generated from a model twice differentiable, both non-differentiable model and the model once differentiable have smaller maximum loglikelihood values than a model twice differentiable. Note that differentiable models perform significantly better than non-differentiable model, for smooth data.

Table 2.10: (Simulation example IV) Sample means and standard deviations of prediction measures for each model (100 cases). True covariance functions are MG($\sigma^2 = 1$, $\alpha = 2000$, $\nu = 0.5$), MI1.8($\sigma^2 = 1$, $\alpha = 2000$, $\nu = 0.5, \epsilon = 100$), and MII1.4($\sigma^2 = 1$, $\alpha = 2000$, $\nu = 0.5$,$\epsilon_1 = 200$, $\epsilon_2 = 100$). For MII1.4, $\hat{\epsilon}_1/\hat{\epsilon}_2$ is displayed.

	Model	RMSPE	MAE	CRPS
MG	MG	0.654(0.179)	0.531(0.153)	0.375(0.100)
	MI1.8	0.654(0.178)	0.530(0.151)	0.375(0.099)
	MII1.4	0.654(0.178)	0.530(0.151)	0.375(0.099)
MI1.8	MG	0.634(0.159)	0.502(0.136)	0.357(0.092)
	MI1.8	0.633(0.155)	0.501(0.131)	0.356(0.089)
	MII1.4	0.633(0.154)	0.503(0.130)	0.357(0.088)
MII1.4	MG	0.565(0.146)	0.446(0.116)	0.317(0.078)
	MI1.8	0.567(0.146)	0.446(0.120)	0.318(0.079)
	MII1.4	0.569(0.145)	0.448(0.119)	0.318(0.077)

2.4 Application

2.4.1 Data and the mean structure

We apply the covariance models to sea level pressure data on a global scale. The data is model outputs from the Beijing Climate Center Climate System Model (http://forecast.bcccsm.cma.gov.cn/). It is a part of the fifth phase of the Coupled Model Intercomparison Project (CMIP5). For more information on CMIP5 archive, see Taylor et al. (2012). We use Boreal summer averages (averaged over June to August, JJA, then averaged over 1971 - 2012). The output values are given on regular grids, and there are 128 longitude points and 64 latitude points; the size of the data is $8,192$. The unit of the sea level pressure is a pascal (Pa). For variance stabilization, we took a square root transform of the data, and it is displayed in Figure 2.5.

Since we are focusing on modeling the covariance structure of data, we subtracted

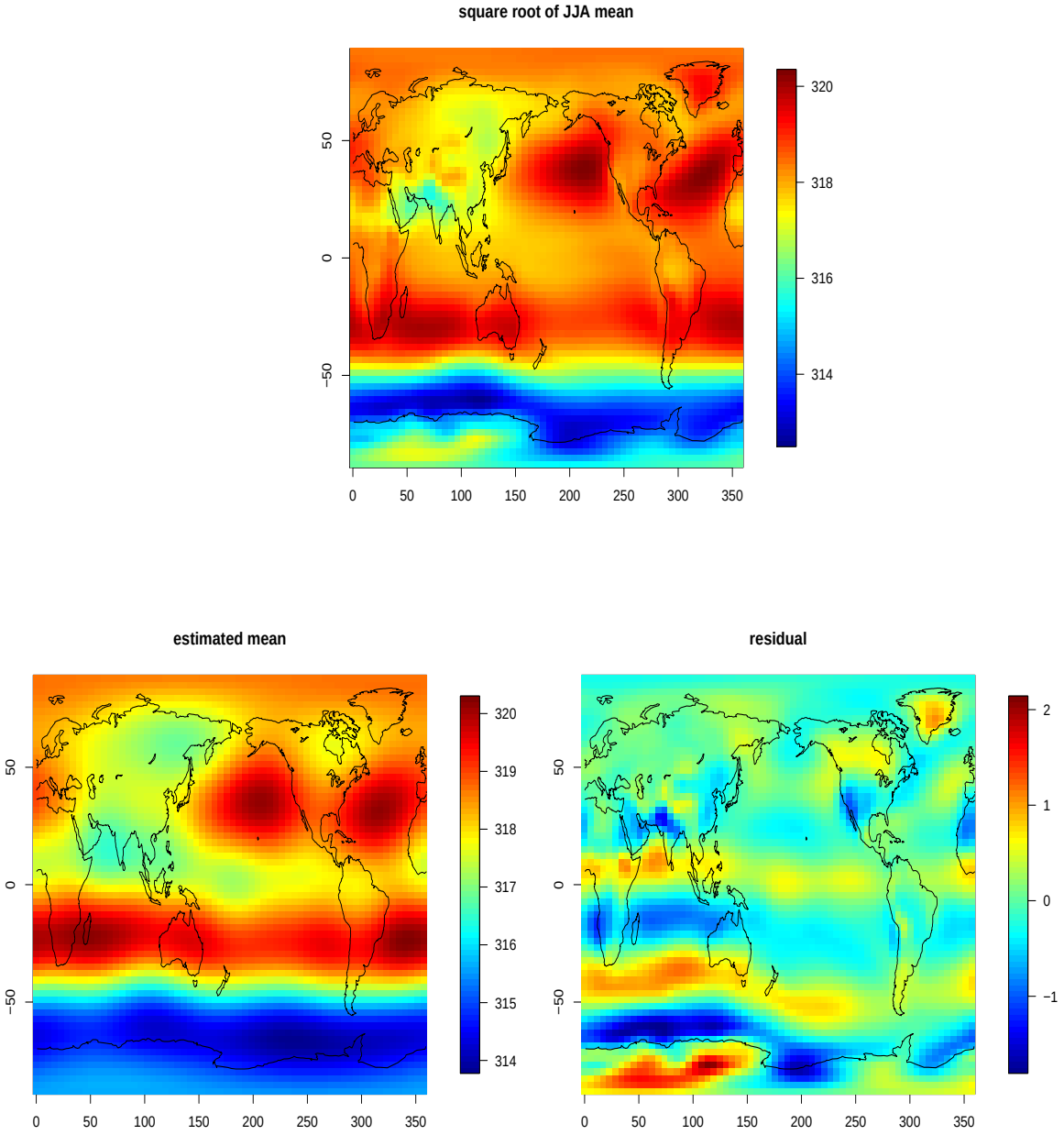

Figure 2.5: (Real data example) Square root of the JJA average of the sea level pressure, its estimated mean structure using spherical harmonics as in (2.11), and the resulting residuals.

the mean structure of data and made it close to mean zero. That is, we filtered out mean structure of data by using spherical harmonics, as in Jun and Stein (2008); we regressed the square root of JJA means with $Y_k^m\{\sin(L), l\}$ for $k = 6$,

$$f(L, l) = \sum_{k=0}^{6} \sum_{m=-k}^{k} f_k^m Y_k^m\{\sin(L), l\}, \tag{2.11}$$

where L and l denote latitude and longitude, respectively. Figure 2.5 shows the estimated mean structure of data and corresponding residuals.

2.4.2 Comparison of covariance models

We compared the following covariance models considered so far, MC, MG, MIx.y, MIIx.y, and NMIx.y, for various x and y values. Moreover, we considered the nonstationary Matérn covariance model (call it, NMC). For NMC, we let ν to be different over the sea and the land. To save computational burden, we randomly selected 1000 locations from the map of residuals. Among 1000 locations, we used 900 randomly chosen locations as training data for parameter estimation, and the remaining 100 data points for validation of prediction. We repeated this process 100 times.

Table 2.11 provides sample means and standard deviations of parameter estimates for each model. For the models MG and MI, estimates for the range parameter are much larger than that of MC. They appear to be associated with smaller estimates of the smoothness parameter and the estimate of ϵ. Table 2.11 also displays sample means and standard deviations of the maximum loglikelihood estimates under each model. The model MII outperforms MI and MG, but it has smaller mean value of maximum loglikelihood than MC.Nonstationary covariance models, NMI and NMC, have larger mean value of maximum loglikelihood than MC. NMC gives somewhat higher loglikelihood value than NMI. For the proposed models, we expect to get

better performances in terms of maximum loglikelihood values by increasing n_c and n_p further.

Table 2.11: (Real data example) Sample means and standard deviations of parameter estimates, maximum loglikelihood values and prediction measures for each model (100 cases). For MII1.4, $\hat{\epsilon}_1/\hat{\epsilon}_2$, for NMI2.6, $\hat{\epsilon}_{sea/land}$, and for NMC, $\hat{\nu}_{sea/land}$ are displayed. For prediction measures, $\text{RMSPE}_{sea/land}$, $\text{MAE}_{sea/land}$, and $\text{CRPS}_{sea/land}$ are displayed.

Model	$\hat{\epsilon}$	$\hat{\sigma}^2$	$\hat{\alpha}$	$\hat{\nu}$	Max.loglik
MC	-	0.28(0.03)	629.72(101.95)	2.47(0.36)	1347.06(51.0)
MG	-	0.35(0.02)	5444.96(247.90)	0.50(0.00)	967.22(16.6)
MI1.8	402.92(33.92)	0.61(0.10)	3948.09(745.10)	0.50(0.00)	1274.96(37.1)
MII1.4	357.19(84.44) 301.23(90.94)	0.41(0.09)	1815.89(476.29)	0.50(0.00)	1319.98(49.4)
NMI2.6	636.83(29.95) 358.13(55.81)	0.41(0.08)	2566.74(560.44)	0.50(0.00)	1363.43(30.2)
NMC	-	0.30(0.04)	1093.85(60.86)	2.35(0.13)	1447.16(37.7)

Table 2.12 shows sample means and standard deviations of prediction measures for each model. As expected, the proposed integrated Matérn models provide smaller RMSPE, MAE, and CRPS values than MG, and their mean value and variability are close to those of MC. Also, MII and NMI perform equally well compared to MC in terms of prediction.

2.5 Conclusion

We have proposed an approach to produce a new class of Matérn-like covariance functions for modeling global data. The resulting covariance functions are positive definite on the surface of a sphere with great circle distance and can deal with smooth processes. There is a straightforward extension of the proposed model for nonstationary data. Through simulation studies and data analysis on a sea level pressure

Table 2.12: (Real data example) Sample means and standard deviations of parameter estimates, maximum loglikelihood values and prediction measures for each model (100 cases). For MII1.4, $\hat{\epsilon}_1/\hat{\epsilon}_2$, for NMI2.6, $\hat{\epsilon}_{sea/land}$, and for NMC, $\hat{\nu}_{sea/land}$ are displayed. For prediction measures, $\text{RMSPE}_{sea/land}$, $\text{MAE}_{sea/land}$, and $\text{CRPS}_{sea/land}$ are displayed.

Model	RMSPE	MAE	CRPS
MC	0.054(0.017)	0.030(0.006)	0.028(0.005)
	0.113(0.042)	0.066(0.020)	0.053(0.017)
MG	0.087(0.023)	0.053(0.009)	0.054(0.006)
	0.153(0.045)	0.098(0.025)	0.080(0.018)
MI1.8	0.058(0.018)	0.032(0.007)	0.031(0.005)
	0.120(0.041)	0.071(0.021)	0.058(0.017)
MII1.4	0.059(0.018)	0.032(0.006)	0.030(0.005)
	0.118(0.043)	0.069(0.022)	0.055(0.018)
NMI2.6	0.058(0.018)	0.033(0.007)	0.030(0.006)
	0.120(0.041)	0.072(0.021)	0.060(0.016)
NMC	0.057(0.016)	0.031(0.006)	0.025(0.005)
	0.124(0.038)	0.077(0.020)	0.062(0.015)

variable, we have shown that new Matérn-like covariance models using great circle distance have substantially better performance in terms of maximum loglikelihood, and prediction for smooth data when we compare with the Matérn covariance model using great circle distance. In several simulation results and real data analysis presented, unless the spatial range was relatively large, the performance of the proposed approach and Matérn covariance model with chordal distance were similar (except for the nonstationary version of the proposed model). Guinness and Fuentes (2013) report similar findings.

It is evident that larger n_c and n_p offer better approximation to the integration, but this will require substantial increase in computational cost. For computational efficiency with large data and large n_c and n_p values, the proposed model can be fitted with recently developed computational techniques, such as covariance tapering

(Furrer et al., 2006; Kaufman et al., 2008) or full scale covariance approximation (Sang and Huang, 2012).

One useful extension for the proposed model is its application to the differential operators approach for nonstationary covariance models on the surface of a sphere (Jun and Stein, 2007, 2008; Jun, 2014). They use chordal distance instead of great circle distance due to mathematical reasons, but adopting the proposed covariance models in their approach will enable them to use great circle distance as a distance metric. Using chordal distance or great circle distance in this case may not make much difference in terms of model fit and spatial prediction in many applications. Nonetheless, as Gneiting (2013) argues, great circle distance is a physically more natural distance metric on the surface of a sphere, and it is worth developing valid covariance models with great circle distance. The proposed model can also be naturally extended to the multivariate case. These are some of the future directions the authors are currently pursuing.

3. COVARIANCE MODELS ON THE SURFACE OF A SPHERE: WHEN DOES IT MATTER?*

3.1 Introduction

In geophysical and environmental sciences, data often come in a global scale. Covariance functions for global data sets need to be positive definite on the surface of a sphere, and the distance computation is important in spatial modeling. For an integer $d \geq 1$, define $\mathcal{S}_r^d = \{\mathbf{x} \in \mathbb{R}^{d+1} : ||\mathbf{x}|| = r\}$ to be a (d-dimensional) sphere with radius r, where $||\mathbf{x}||$ is the Euclidean norm of $\mathbf{x} \in \mathbb{R}^{d+1}$. We also define the great circle distance on $\mathcal{S}_r^d$ by $\theta(\mathbf{x}, \mathbf{y}) = r \times \arccos(\langle \mathbf{x}, \mathbf{y} \rangle)$ where $\langle \cdot, \cdot \rangle$ denotes the usual inner product on $\mathbb{R}^{d+1}$.

We consider the surface of the Earth as the spatial domain. Let $\mathcal{S}_R^2$ denote the surface of the Earth, where R denotes the Earth's radius (we approximate the Earth as a perfect sphere). The chordal distance between the two points, $\mathbf{s}_1$=(L_1,l_1) and $\mathbf{s}_2$=(L_2,l_2), on $\mathcal{S}_R^2$ (L and l denote latitude and longitude, respectively) is given by

$$\text{ch}(\mathbf{s}_1, \mathbf{s}_2) = 2R[\sin^2\{(L_1 - L_2)/2\} + \cos L_1 \cos L_2 \sin^2\{(l_1 - l_2)/2\}]^{1/2}.$$

The great circle distance between the two locations then is given by $\theta = \theta(\mathbf{s}_1, \mathbf{s}_2) = 2R \times \arcsin\{\text{ch}(\mathbf{s}_1, \mathbf{s}_2)/(2R)\}$. The chordal distance is simply the Euclidean distance penetrating the spatial domain on the surface of the Earth and producing a straight-line approximation to the great circle distance (Banerjee, 2005).

Yadrenko (1983) pointed out that any covariance function in $\mathbb{R}^3$ can be considered as a covariance function for processes on $\mathcal{S}_r^2$ using the chordal (i.e. Euclidean)

*Reprinted from *Stat*, 4, Jeong, J. and Jun, M., Covariance models on the surface of a sphere: when does it matter?, 167-182, 2015, with permission from John Wiley & Sons, Inc.

distance. This construction can provide a rich class of covariance functions on $\mathcal{S}_r^2$ (Gneiting, 1999). As argued by Gneiting (2013), the great circle distance is a physically most natural distance metric for processes on a sphere. However, literature on covariance modeling using the great circle distance on the surface of a sphere is scarce because of its mathematical challenge. Some efforts have been made in examining the validity of several parametric covariance functions on the surface of a sphere (Huang et al., 2011; Gneiting, 2013) and in developing valid parametric covariance functions with the great circle distance from various constructional approaches (Du and Ma, 2012; Du et al., 2013; Gneiting, 2013; Guinness and Fuentes, 2013; Jeong and Jun, 2015).

Although Huang et al. (2011) and Gneiting (2013) studied validity of covariance functions defined with either the great circle distance or the Euclidean distance in detail, the impact upon parameter estimation and prediction has not been studied well. According to Banerjee (2005), careless formulation of distances can lead to poor prediction with the wrong estimation of the spatial range. Note that this study considered the Matérn covariance model using the great circle distance, which may not be positive definite on the surface of a sphere, unless the smoothness parameter, ν, is $\nu \in (0, 0.5]$ (Miller and Samko, 2001; Gneiting, 2013).

In this Section, we consider several positive definite functions on $\mathbb{R}^{d+1}$ and $\mathcal{S}_r^d$ for $d = 1, 2$ and compare them in simulation studies and real data applications. The rest of the Section is organized as follows. In Section 3.2, we discuss some characteristics of covariance functions on a sphere. Then we present two simulation studies on $\mathcal{S}_1^1$ and $\mathcal{S}_R^2$ in Section 3.3. Section 3.4 illustrates real application results to geopotential height data set. Finally, Section 3.5 concludes the Section with discussion.

3.2 Characteristics of covariance functions on a sphere

We first review some known results on covariance functions in the Euclidean space as well as those on the surface of a sphere. A function $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is called positive definite if

$$\sum_{i,j=1}^{n} c_i c_j f(\mathbf{x}_i, \mathbf{x}_j) \geq 0 \tag{3.1}$$

for all finite $n \in \mathbb{N}$, all distinct points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$, and all real $c_1, \ldots, c_n$. A function is strictly positive definite when the inequality in (3.1) is strict (unless $c_1 = c_2 = \cdots = c_n = 0$). For a real random field Z in $\mathbb{R}^d$ with $E\{Z(\mathbf{x})\}^2 < \infty$ for all $\mathbf{x} \in \mathbb{R}^d$, the covariance function $K(\mathbf{x}, \mathbf{y}) = \text{cov}\{Z(\mathbf{x}), Z(\mathbf{y})\}$ must satisfy the condition in (1). The random field Z is called weakly stationary if its mean function is constant, it has finite second moments, and its covariance function can be written as $\text{cov}\{Z(\mathbf{x}), Z(\mathbf{y})\} = K(\mathbf{x} - \mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, and a positive definite function K, that is. the covariance function of Z, depends on $\mathbf{x}$ and $\mathbf{y}$ only through $\mathbf{x}-\mathbf{y}$. Furthermore, if its covariance function satisfies $\text{cov}\{Z(\mathbf{x}), Z(\mathbf{y})\} = \varphi(||\mathbf{x}-\mathbf{y}||)$ for a positive definite function φ, then the random field Z is weakly isotropic. An isotropic property for processes in $\mathbb{R}^d$ can be thought as an invariance property under translation and rotation (Stein, 1999).

The covariance function of a random field and the smoothness of its realization are related to mean square properties of the random field. The random field Z is called mean square continuous at $\mathbf{x}$ if

$$E\{Z(\mathbf{y}) - Z(\mathbf{x})\}^2 \to 0 \quad \text{as } \mathbf{y} \to \mathbf{x}.$$

For weakly stationary random field, mean square continuity is equivalent to the fact that the covariance function is continuous at the origin, but it does not im-

ply continuity of its realization (Stein, 1999). Moreover, a random field Z on $\mathbb{R}$ with finite second moments is mean square differentiable at t if there exists $Z'(t) = \lim_{n\to\infty}\{Z(t+h_n) - Z(t)\}/h_n$ in L^2 for sequences $h_n \to 0$. The smoothness of a random field can be determined through the number of mean square derivatives. Gneiting (2013) defined the class of Φ_d with the correlation functions of mean square continuous, stationary and isotropic random fields in $\mathbb{R}^d$. Every positive definite function $\varphi : [0,\infty) \to \mathbb{R}$ with $\varphi(0) = 1$ is the correlation of an isotropic process, and the members of Φ_2 and Φ_3 are the cornerstones for covariance models for spatial data in a planar domain (Gneiting, 2013). An isotropic property on a sphere means that the covariance function depends on distance only. That is, a random field Z on $\mathcal{S}_r^d$ is called isotropic if its covariance function satisfies $\text{cov}\{Z(\mathbf{x}), Z(\mathbf{y})\} = \psi(\theta(\mathbf{x},\mathbf{y}))$ for all $\mathbf{x},\mathbf{y} \in \mathcal{S}_r^d$. We then similarly define Ψ_d, the class of continuous, isotropic covariance functions $\psi : [0, \pi \times r] \to \mathbb{R}$ on $\mathcal{S}_r^d$.

Because a sphere can be viewed as a subset of the Euclidean space, valid covariance functions on $\mathbb{R}^{d+1} \times \mathbb{R}^{d+1}$ can be restricted to $\mathcal{S}_r^d \times \mathcal{S}_r^d$ when the Euclidean distance is used (equivalently, the chordal distance on $\mathcal{S}_r^d$). Yadrenko (1983) and Yaglom (1987) pointed out that if φ is a member of the class Φ_{d+1}, then the function $\varphi[2r\sin\{\theta/(2r)\}]$, with the Euclidean distance expressed in terms of great circle distance as $2r\sin\{\theta(\mathbf{x},\mathbf{y})/(2r)\}$, belongs to the class Ψ_d. Because there are various positive definite functions, including the Matérn class and the generalized Cauchy families (Stein, 1999; Gneiting, 2013) that are isotropic covariance functions for processes in $\mathbb{R}^3$, this mapping from $\varphi \in \Phi_3$ to $\psi \in \Psi_2$ provides a useful way to construct a rich parametric class of isotropic covariance functions on $\mathcal{S}_r^2$. This mapping preserves the interpretation of parameters such as scale, range, smoothness and fractal index (Gneiting, 2013).

It has been reported in the literature (e.g. Guinness and Fuentes (2013), Jeong

and Jun (2015)) that when Matérn class with the Euclidean distance and that with the great circle distance are compared in terms of model fit and prediction, often Matérn model with the Euclidean distance performs better. This may be due to the restriction on the smoothness parameter for the Matérn class with the great circle distance. Jeong and Jun (2015) proposed a method to overcome such limitation on the smoothness parameter for the Matérn class with the great circle distance, but they found that the Matérn class with the Euclidean distance is equivalent or often better compared with the covariance models specifically developed for processes on the sphere. Our goal in this Section is to study cases that are not previously considered in the literature and to explore cases where there are significant differences (improvements) of covariance models defined on the sphere as opposed to the covariance models projected from the Euclidean space.

We focus on the fact that there are some fundamental differences between covariance models originally defined on the surface of a sphere and those in the Euclidean space. For instance, there exists a lower bound on isotropic correlation function in the Euclidean space. A function φ is an isotropic correlation function in $\mathbb{R}^d$ if and only if it has the form, $\varphi(t) = \int_0^\infty \Lambda_d(tu)dG(u)$, where $\int_0^\infty dG(u) = 1$ and G is non-decreasing. Note that $\Lambda_d(r) = 2^{(d-2)/2}\Gamma(d/2)r^{-(d-2)/2}J_{(d-2)/2}(r)$ where J is a Bessel function. Then, for all t,

$$\varphi(t) \geq \inf_{s\geq 0}\Lambda_d(s) \tag{3.2}$$

(Stein, 1999). This implies that valid correlation functions on $\mathcal{S}_r^2$ constructed through the mapping described earlier from $\varphi \in \Phi_3$ cannot have values less than $\inf_{s\geq 0} s^{-1}\sin s = -0.218$. In particular, the Matérn class yields non-negative correlations only. Although the importance of the Matérn family is highlighted by Stein (1999) because of its flexibility with regard to the local behavior of the processes, it might not be

appropriate in applications where there is a significantly negative spatial correlation. In fact, many of the isotropic covariance functions in $\mathbb{R}^d$ used in the literature take non-negative values only.

We also focus on the fact that on the sphere, the correlation between two points with large distance apart may not necessarily be small (compared with the correlation between nearby two points). In fact, if there is a wave traveling around the sphere, two points nearly maximum possible distance apart may be perfectly positively (or negatively) correlated, which cannot happen in the Euclidean space.

We now list several parametric classes of covariance functions defined on the surface of a sphere or defined originally in the Euclidean space (then can be used through the projection). Some of the models are used in the simulation and data examples. We only consider isotropic covariance models on a sphere. Functions in the class Ψ_d are characterized in terms of an infinite sum of Gegenbauer polynomials with non-negative coefficients and cosine of the great circle distance (Schoenberg, 1942; Gneiting, 2013). For $d \geq 1$, the class Ψ_d consists of the functions of the form

$$\psi(\theta) = \sum_{n=0}^{\infty} b_{n,d} C_n^{(d-1)/2}(\cos(\theta/r))/C_n^{(d-1)/2}, \quad \theta \in [0, \pi \times r],$$

with non-negative coefficients $b_{n,d}$ such that $\sum_{n=0}^{\infty} b_{n,d} = 1$ and the Gegenbauer polynomial of degree n, $C_n^{(d-1)/2}$ (Schoenberg, 1942; Chen et al., 2003). Moreover, the class Ψ_∞ consists of the functions with the following form

$$\psi(\theta) = \sum_{n=0}^{\infty} b_n (\cos(\theta/r))^n, \quad \theta \in [0, \pi \times r], \tag{3.3}$$

with non-negative coefficients b_n such that $\sum_{n=0}^{\infty} b_n = 1$. The infinite sum is strictly positive definite on $\mathcal{S}_r^d$ when the coefficients $b_{n,d}$ and b_n are strictly positive for

infinitely many odd and infinitely many even integers n, and only a few closed forms, such as the multiquadratic family, for such infinite sums are known in general. The multiquadratic covariance function (Gneiting, 2013) is defined by

$$\psi(\theta) = \sigma^2(1-\tau)^{2c}/\{1+\tau^2-2\tau\cos(\theta/r)\}^c, \quad \theta \in [0, \pi \times r]$$

from a standard Taylor series of (3.3), when $\sigma^2 > 0$, $c > 0$, and $\tau \in (0,1)$. The Matérn class, given as

$$\varphi(t) = \sigma^2 2^{\nu-1}\Gamma(\nu)^{-1}(t/\alpha)^\nu \mathcal{K}_\nu(t/\alpha), \quad t \geq 0, \tag{3.4}$$

where the parameters, $\sigma^2, \alpha, \nu > 0$, are marginal variance, spatial range and smoothness parameters, respectively, is positive definite in $\mathbb{R}^d$ for any $d \in \mathbb{N}$ with the Euclidean distance (Stein, 1999). Here, $\mathcal{K}_\nu$ is the modified Bessel function of the second kind of order ν.

Gneiting (2013) showed that completely monotone functions (that have derivatives $\varphi^{(k)}$ of all orders with $(-1)^k\varphi^{(k)}(t) \geq 0$ for all non-negative integers k and all positive t) including the power exponential, Matérn and generalized Cauchy families are positive definite, through the restriction of a function $\varphi : [0,\infty) \to \mathbb{R}$ to the interval $[0, \pi \times r]$: $\psi = \varphi_{[0,\pi\times r]}$ under applicable conditions, on $\mathcal{S}_r^d$ of any dimension. One necessary condition of the membership in the class Ψ_d is that either the fractal index or the smoothness parameter requires to satisfy $\beta \in (0,1]$ or $\nu \in (0, 0.5]$, respectively. The powered exponential family defined by

$$\psi(\theta) = \sigma^2 \exp[-\{\theta/(cr)\}^\beta], \quad \theta \in [0, \pi \times r]$$

where $\sigma^2 > 0$, $c > 0$ is valid on any dimensional $\mathcal{S}_r^d$ if $\beta \in (0,1]$, and the Matérn

family requires $\nu \in (0, 0.5]$ similarly. Jeong and Jun (2015) compare Matérn class with great circle distance to that the Euclidean distance for spatial data on the surface of a sphere.

Compactly supported members of the class Φ_3 may be valid on $\mathcal{S}_r^2$ through $\psi = \varphi_{[0,\pi\times r]}$ Gneiting (2013). The spherical and the Wendland's functions on Φ_3, remain valid with direct substitution of the Euclidean distance by the great circle distance on Ψ_d, $d = 1, 2, 3$. In particular, the C^4-Wendland covariance function (Wendland, 1995), defined as

$$\psi(\theta) = \sigma^2\{1 + (\theta\tau)/(cr) + \theta^2(\tau^2 - 1)/(3c^2r^2)\}\{1 - \theta/(cr)\}_+^\tau, \quad \theta \in [0, \pi \times r], \quad (3.5)$$

where $c \in (0, \pi]$ is a support parameter and $\tau \geq 6$ is a shape parameter, has four derivatives at the origin and thus may be suitable for smooth data on $\mathcal{S}_r^d$.

The following covariance function (defined in the Euclidean space) models a hole effect,

$$\varphi(t) = \sigma^2(\alpha/t)\sin(t/\alpha), \quad t > 0, \quad (3.6)$$

with $\varphi(0) = \sigma^2$, and it is called the wave covariance function. This function is able to deal with the situation where the correlation between two points far apart is bigger (in magnitude) than the correlation between two points closer, or there is an oscillating pattern in the correlation function. This belongs to the class Φ_d for $d = 1, 2, 3$, but it is not valid on $\mathcal{S}_r^2$ (Huang et al., 2011). Thus, this function can only be applied to the data on the sphere through the projection from the Euclidean space.

Schoenberg (1942) noted that the class Ψ_d enjoys the useful closure properties. The class Ψ_d is convex, closed under products and closed under limits, which are

continuous. For example, if $\psi_1(\theta), \psi_2(\theta) \in \Psi_d$, then $\lambda\psi_1(\theta) + (1-\lambda)\psi_2(\theta) \in \Psi_d$ for every $\lambda \in [0,1]$ and $\psi_1(\theta) \times \psi_2(\theta) \in \Psi_d$. Moreover, if $\psi_n(\theta) \in \Psi_d$, $\psi_n(\theta) \to \psi(\theta)$ as $n \to \infty$, and $\psi(\theta)$ is continuous, then also $\psi(\theta) \in \Psi_d$. These closure properties of the class Ψ_d offer additional flexibility to model negative correlations. If we use convex sums or products of parametric families of correlation functions on $\mathcal{S}_r^d$ in terms of the great circle distance with a Legendre function of the form $\psi(\theta) = C_n^{(d-1)/2}\cos(\theta/r)/C_n^{(d-1)/2}$, including the case $\cos(\theta/r)$ for $d = 1, 2, 3$, we can easily model negative correlations (Gneiting, 2013).

The sine-power covariance function is defined by

$$\psi(\theta) = \sigma^2\{1 - (\sin\tfrac{\theta}{2r})^\beta\}, \quad \theta \in [0, \pi \times r], \tag{3.7}$$

where the parameter $\beta \in (0, 2]$ corresponds to the fractal index and controls the smoothness of the process. This covariance function operates directly on a sphere (belongs to the class Ψ_d for all dimensions). On the other hand, the cosine function, $\varphi(t) = \cos(nt)$, where $t \geq 0$, belongs to the class Φ_1 only for any value of $n > 0$. On $\mathcal{S}_r^d$, $\psi(\theta) = \cos(n\theta/r)$ for $\theta \in [0, \pi \times r]$, is non-strictly positive definite for all dimensions when $n = 1$. For non-integer values of $n > 0$, $\psi(\theta)$ is not valid on $\mathcal{S}_r^1$ and for integer values of $n \geq 2$, it is non-strictly positive definite only on $\mathcal{S}_r^1$ but not on $\mathcal{S}_r^2$ (Gneiting, 2013).

When there is no significantly negative correlation, and when the spatial range is small, we do not expect that using the great circle distance or the Euclidean distance may make a significant difference. Moreover, when the prediction location is surrounded by enough number of estimation locations, we do not expect significant difference between models defined on the sphere and models defined in the Euclidean space (and used through projection). Therefore, for the rest of the Section, we mainly

consider the situations where either spatial range is large, or when there are not many estimation locations close to the prediction locations. Such situation may arise in real application, for example, when we are interested in predicting over the ocean with most observations on the land or when we are interested in predicting near the poles with not many observations near the poles. We also consider cases where there are significantly negative correlations at large distance lags.

3.3 Simulation studies

We present two simulation studies on $\mathcal{S}_1^1$ and $\mathcal{S}_R^2$. In the first example (Section 3.3.1), the truth is generated from an exponential covariance model with the great circle distance. In the second example (Section 3.3.2), the truth is generated from oscillating Matérn covariance model implemented in `R-INLA` package (Martins et al., 2013).

3.3.1 Example on $\mathcal{S}_1^1$

We simulated mean zero Gaussian random fields on $\mathcal{S}_1^1$ using exponential covariance models with various range parameters using the great circle distance. The exponential covariance function on $\mathcal{S}_r^d$ is defined by

$$\psi(\theta) = \sigma^2 \exp(-\theta/\alpha), \quad \theta \in [0, \pi \times r],$$

where $\sigma^2 > 0$ is a variance and $\alpha > 0$ is a spatial range parameter. We then compared fitted results using the true covariance model as well as exponential covariance model with the Euclidean distance, that is, $\varphi(t) = \sigma^2 \exp(-t/\alpha)$, $t > 0$. Exponential covariance functions belong to both Φ_2 and Ψ_1, so they are valid on the surface of a sphere regardless of distance. We set the marginal variance $\sigma^2 = 1$ and considered the various spatial ranges $\alpha = 2\pi, 1.5\pi, \pi, \pi/1.5, \pi/2, \pi/4$. We randomly selected

100 locations (angles) for parameter estimation from $(\pi/2, 3\pi/2)$ on a unit circle and 10 fixed and equally spaced locations for prediction from $[0, \pi/2)$ as in Figure 3.1. We compared the covariance models in terms of prediction using the root mean square error (RMSE) as well as the two popular scoring rules (Gneiting and Raftery, 2007), the mean absolute error (MAE) and the continuous ranked probability score (CRPS). We repeated this experiment 100 times: the sampling locations are different each time, and the prediction locations are fixed.

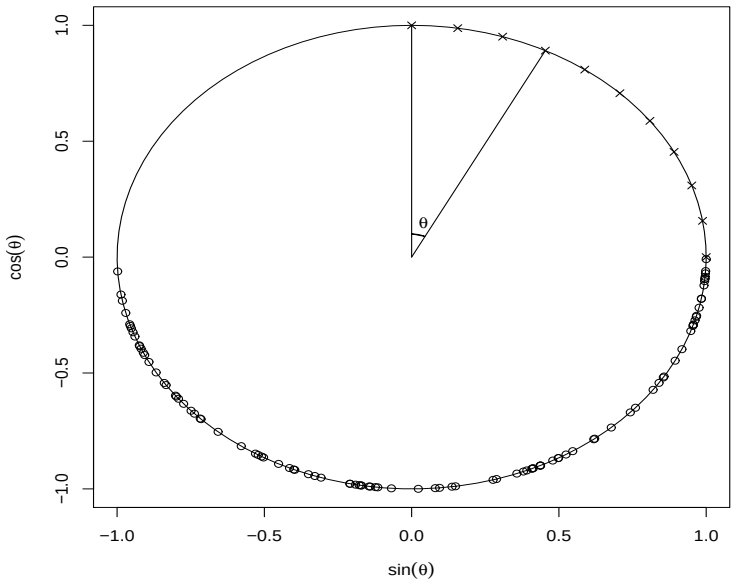

Figure 3.1: (Simulation example on $\mathcal{S}_1^1$) A realization of sampling locations ($\circ$) and prediction locations ($\times$).

Figure 3.2 presents MAE and CRPS from the two models, the exponential models using great circle and chordal distances, displayed against 10 fixed prediction locations, for the true spatial range values are 2π and $\pi/4$. For the larger spatial

range, $\alpha = 2\pi$, we observe that the exponential model using the great circle distance performed significantly better than that using the chordal distance. Except for prediction locations that are relatively close to sampling locations, there are considerable differences between two models in prediction errors. On the other hand, for the smaller spatial range, $\alpha = \pi/4$, there is no significant difference between two models in both prediction errors, which agrees with findings of Guinness and Fuentes (2013) and Jeong and Jun (2015).

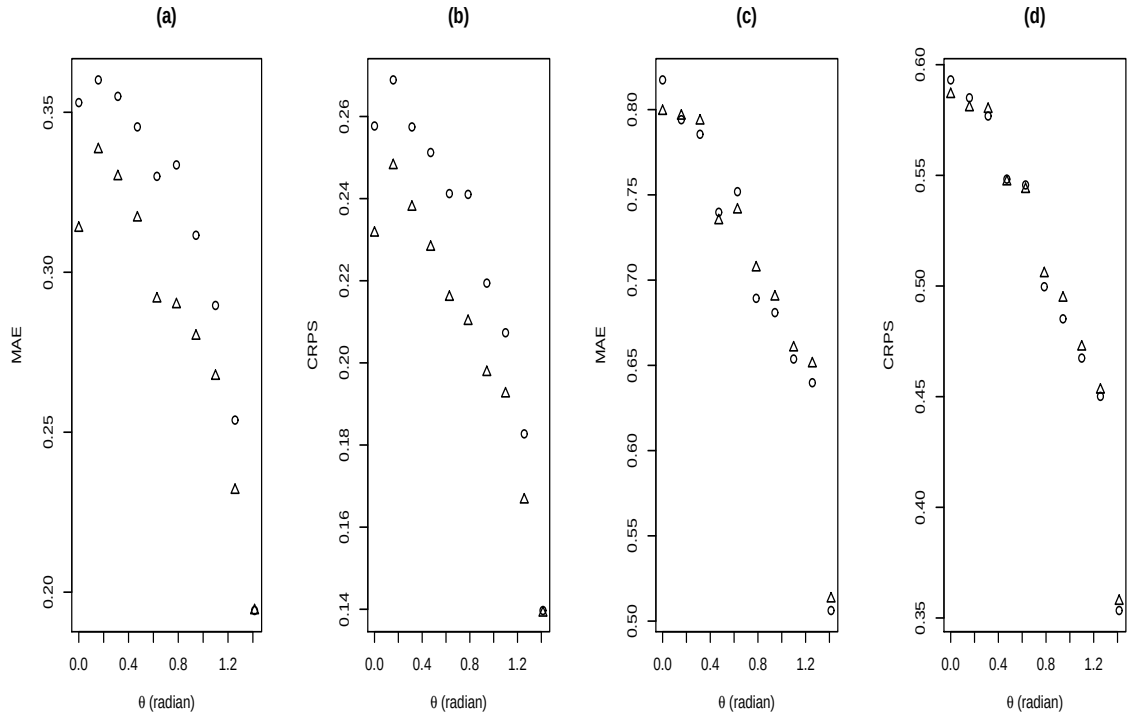

Figure 3.2: (Simulation example on $\mathcal{S}_1^1$) MAE (a) and mean CRPS (b) averaged over 100 replications from the two models displayed against prediction locations (θ) when the true spatial range $\alpha = 2\pi$. (c) and (d) are the same as (a) and (b), except that $\alpha = \pi/4$. Triangles and circles represent the values of prediction errors for the exponential models using great circle and chordal distances, respectively.

3.3.2 Example on $\mathcal{S}_R^2$

We consider mean zero Gaussian random fields on the surface of Earth (with radius $R = 6371$ (km)) with a Matérn covariance function from oscillating stochastic partial differential equations (SPDE) models with the `spde1` class from the `R-INLA` version 0.0-1413638221. We set $\sigma = 1$, $\kappa = 0.5$, $\tau = \kappa \times \sigma / \sqrt{4\pi}$, $\nu = 2$ and $\theta_{osc} = 0.3$. Here κ is the spatial scale parameter, τ controls the variance with σ, ν controls the smoothness of the process and θ_{osc} controls the strength of oscillation. We subtracted the constant mean (average over all locations) to have mean zero residual fields as in Figure 3.3.

There are 128 longitude points and 64 latitude points, and the size of the data is 8192. We randomly selected 300 locations where values are smaller than 0 for parameter estimation and 100 locations where values are larger than 1 for prediction. We repeated this procedure 100 times: all locations are different each time. It is clear from Figure 3.3 that values at large distance lags are negatively correlated. We compared fitted results from a Matérn covariance model with the Euclidean distance (MC) to those from a Matérn covariance model with the great circle distance (MG). Moreover, to deal with negative correlations, we use convex sum of valid covariance functions with the great circle distance (C):

$$\psi(\theta) = \sigma^2[\lambda\{1 - (\sin\tfrac{\theta}{2R})^\beta\} + (1-\lambda)\cos(\theta/R)],\ \beta \in (0,2] \text{ and } \lambda \in (0,1),$$

We also considered the hole-effect model with the Euclidean distance (H) defined in (3.6), a model defined with the Euclidean distance, for a comparison.

Tables 3.1 and 3.2 contain results of parameter estimation and prediction for the various models considered. From Table 3.2, we observe that both MC and MG have large estimates of the spatial range parameter and MG has smaller prediction errors

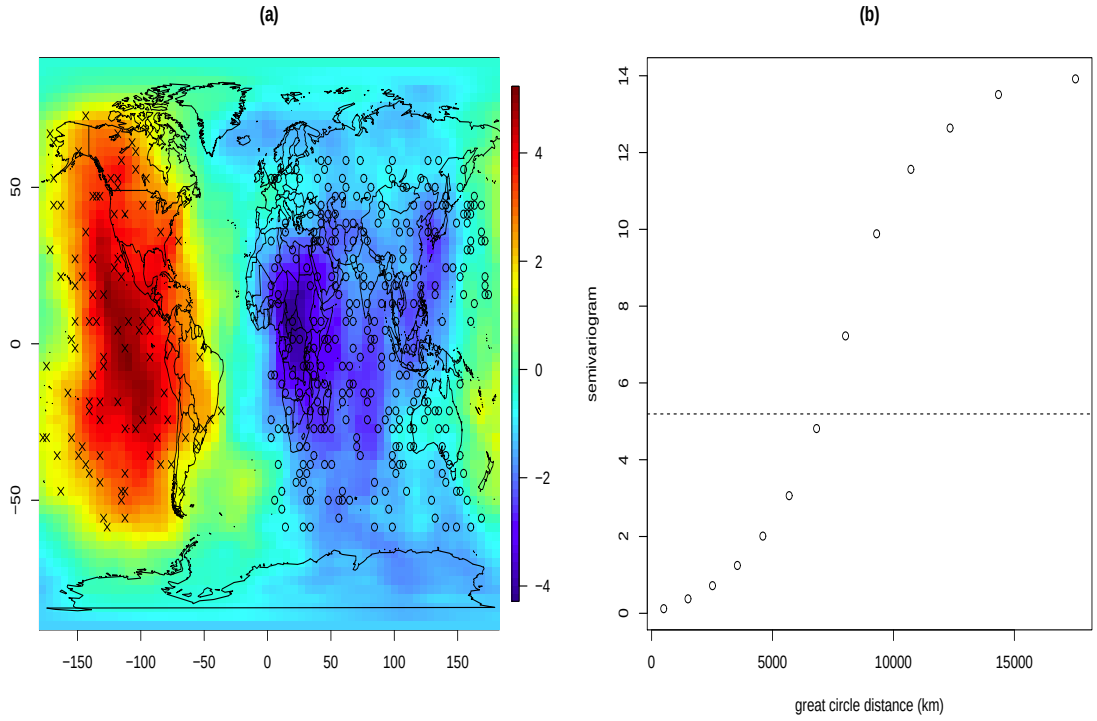

Figure 3.3: (Simulation example on $\mathcal{S}_R^2$) (a) A realization of residual fields. (a) shows sampling locations ($\circ$) and prediction locations ($\times$). (b) Empirical semivariogram values for selected locations versus the great circle distance. For (b), dotted line represents sample variance.

Table 3.1: (Simulation example on $\mathcal{S}_R^2$) Sample means and standard deviations of parameter estimates and maximum log-likelihood values for each model (100 cases). For model C, $\lambda \in (0, 1)$ is a weighting parameter.

Model	$\hat{\sigma}^2$	$\hat{\alpha}$	$\hat{\nu}$	Max.loglik
MC	6.654(1.237)	7918.459(1669.169)	1.037(0.054)	356.505(11.217)
MG	3.260(0.448)	49423.925(7016.223)	0.500(0.000)	298.233(6.923)
C	26.033(7.163)	0.999(0.006)	1.820(0.038)	354.384(10.533)
H	7.496(1.711)	58.499(56.164)	·	-430.486(38.393)

Table 3.2: (Simulation example on $\mathcal{S}_R^2$) Sample means and standard deviations of prediction errors for each model (100 cases). For model C, $\lambda \in (0, 1)$ is a weighting parameter.

Model	RMSPE	MAE	CRPS
MC	1.123(0.179)	0.903(0.177)	0.639(0.096)
MG	0.975(0.085)	0.749(0.065)	0.573(0.058)
C	0.874(0.126)	0.717(0.118)	0.514(0.055)
H	1.961(0.369)	1.751(0.337)	1.166(0.213)

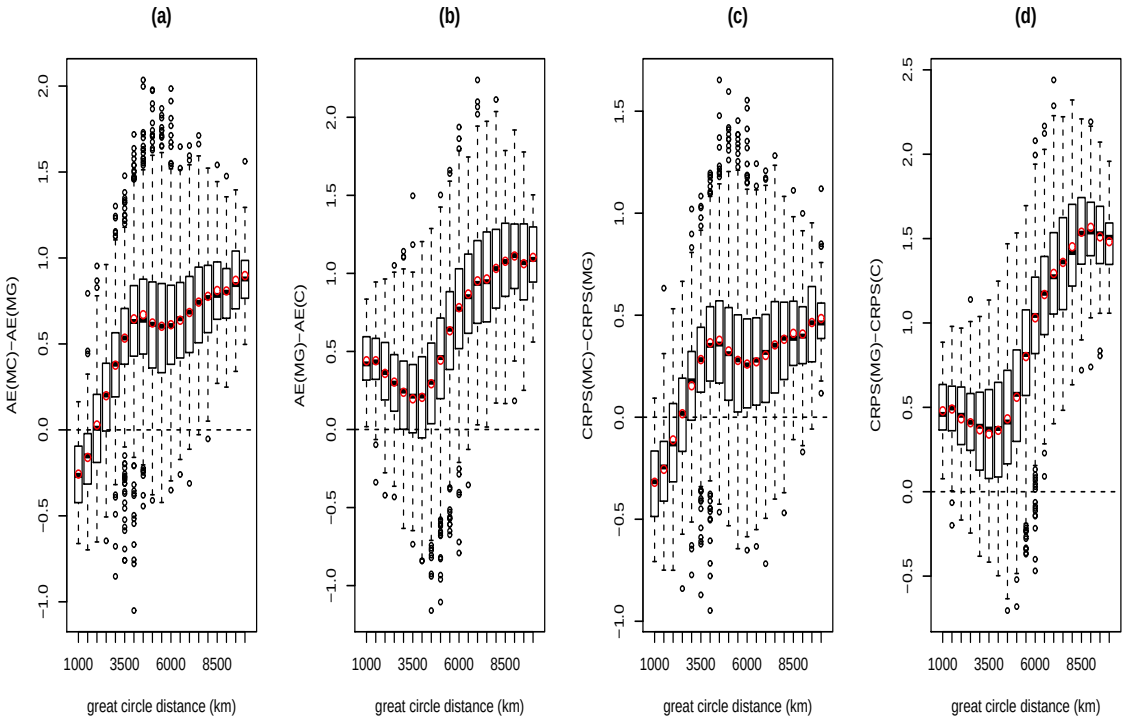

Figure 3.4: (Simulation example on $\mathcal{S}_R^2$) Boxplots of differences of absolute error (AE) (a) and CRPS (c) values from MC and MG, displayed against minimum great circle distance between a prediction location and its nearest sampling location. (b) and (d) The same as (a) and (c), except that selected models are MG and C. For (a)-(d), red circles represent average values in each bin.

than MC. Although MC has larger maximum log-likelihood values than MG, MC leads to poor prediction, possibly because of the large estimate of the spatial range. Note that the best model in terms of prediction errors is C. For C, the estimate of λ is close to 1, and resulting estimated model is dominated by the sine-power model. Nevertheless, its correlation function allows much smaller correlation values for large distance lag, compared with the models for the Euclidean space. On the other hand, for H, although it allows negative correlations unlike Matérn covariance model, it resulted in poor model fit and spatial prediction. From Figure 3.4, we observe clear differences of prediction errors between models using the great circle and Euclidean distances when the prediction locations are relatively far away from their nearest sampling locations. Overall, C outperforms MC and MG in prediction.

3.4 Application

3.4.1 Data and the mean structure

We consider geopotential height data on a global scale. The geopotential height approximates the actual height of a surface pressure at a certain level above mean sea level. It is an adjustment to geometric height using the variation of gravity with elevation and latitude. The study of the geopotential height might be important in learning abnormal weather phenomena. According to Hafez and Almazroui (2014), the geopotential height at level 500 hPa plays a dominant role in controlling weather and climate conditions. Moreover, it became evident that the variability of global geopotential height is clearly impacted by global warming and climatic indices over the last several decades (Marshall, 2002; Zhu et al., 2002; Hafez, 2012; Hafez and Almazroui, 2014).

Data sets are obtained at level 500 hPa from the reanalysis project of National Center for Environmental Prediction (NCEP) and National Center for Atmo-

spheric Research (NCAR) (http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html). The NCEP/NCAR reanalysis 1 project uses a state-of-the-art analysis/forecast system to perform data assimilation using past data from 1948 to the present (see Kalnay et al. (1996) for more detailed information). The output values are given on regular grids, and there are 144 longitude points and 73 latitude points (the size of the data is 10,512). We used Boreal summer geopotential height in the northern hemisphere (June, July, and August; JJA), and for each grid, we computed a pointwise mean as the average over 2014. The unit for the geopotential height is meter and the unit for distance is kilometer. For variance stabilization, we took a square root transform of the data.

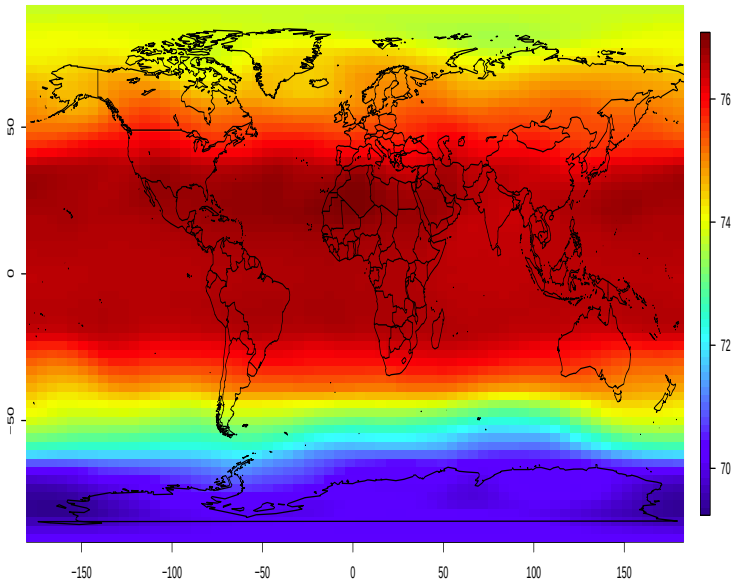

Figure 3.5: (Data example) Square root of the geopotential height at level 500 hPa ($\sqrt{m}$).

We decompose the data into its mean structure (large-scale variation) and the residual (for small-scale spatial variation). Figure 3.5 suggests clear large scale spatial structure depending on latitude. Thus, we modeled the mean structure through simple harmonic regression:

$$m(L) = a_0 + a_1 \cos(L \times \pi/90^\circ) + a_2 \sin(L \times \pi/90^\circ). \tag{3.8}$$

We considered two cases: one is with a constant mean (that is, $a_1 = a_2 = 0$ in (3.8)), and the other is given by (3.8). In both cases, we first estimate the mean structure using regression and then work with the residual to fit the covariance structure.

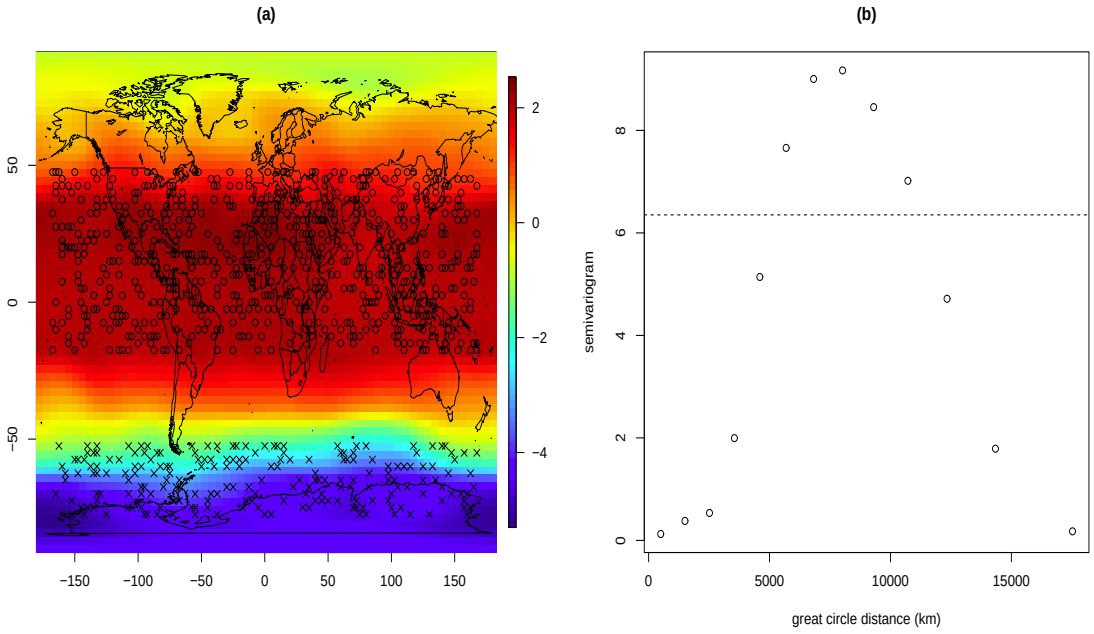

Figure 3.6: (Data example I) (a) A realization of residual fields after subtracting the constant mean. (a) shows sampling locations (∘) and prediction locations (×). (b) Empirical semivariogram values for selected locations versus the great circle distance. For (b), dotted line represents sample variance.

3.4.2 Example I: horizontal directional sampling design for prediction

Figure 3.6 shows the map of residual after subtracting the constant mean and the empirical semivariogram of residuals. The semivariogram clearly shows negative covariances for large distances. To save computational burden, we randomly selected 600 locations near the red region for parameter estimation and selected 200 locations over the blue region for prediction as in Figure 3.6. We repeated this process 100 times, and for each time, all locations are randomly sampled and thus different.

We compared the covariance models, MC, MG and the convex sum model, C, with C^4-Wendland covariance function (3.5) and cosine function. Although MC obtains the largest maximum log-likelihood value as in Table 3.3, it leads to poor prediction compared with C. This is expected as the Matérn model is not able to produce negative correlations. Similarly to the simulation example of Section 3.3.2, C is the best model in terms of prediction. As shown in Table 3.4, MG gives poor model fit and prediction.

Table 3.3: (Data example I - constant mean) Sample means and standard deviations of parameter estimates and maximum log-likelihood values for each model (100 cases). For model C, $c \in (0, \pi]$ is a support parameter and $\tau \geq 6$ is a shape parameter.

Model	$\hat{\sigma}^2$	$\hat{\alpha}$ or $\hat{\lambda}$	$\hat{\nu}$ or $\hat{c}$	$\hat{\tau}$	Max.loglik
MC	1.41(0.09)	2128.23(114.51)	2.61(0.07)	·	2167.40(20.8)
MG	0.80(0.12)	136377.19(18394.33)	0.50(0.00)	·	1413.05(11.1)
C	2.24(0.38)	0.81(0.11)	2.93(0.08)	7.83(0.30)	2160.82(20.4)

Figure 3.7 shows boxplots of differences of absolute error (AE) and CRPS values between MC and C, displayed against minimum great circle distance between a

Table 3.4: (Data example I - constant mean) Sample means and standard deviations of prediction errors for each model (100 cases). For model C, $c \in (0, \pi]$ is a support parameter and $\tau \geq 6$ is a shape parameter.

Model	RMSPE	MAE	CRPS
MC	3.972(0.137)	3.853(0.142)	3.427(0.145)
MG	6.939(0.086)	6.840(0.090)	6.767(0.089)
C	3.292(0.186)	3.168(0.193)	2.650(0.196)

prediction location and its nearest sampling location. For all distances, C outperforms MC. Moreover, the differences of two prediction errors between two models increase as minimum distances between prediction locations and their nearest sampling locations increase. On the other hand, Figure 3.8 presents residual fields after removing mean structure by using simple harmonic regression depending on latitude as in (3.8). We fitted C^4-Wendland covariance functions with the Euclidean distance (WC) and the great circle distance (WG) in addition to the covariance models considered previously with constant mean structure. For WC, the covariance function is defined by

$$\varphi(t) = \sigma^2\{1 + (t\tau)/(cR) + t^2(\tau^2 - 1)/(3c^2R^2)\}\{1 - t/(cR)\}_+^\tau,$$

where $\tau \geq 6$, $c > 0$, and $t > 0$. The sampling and prediction locations remained the same as in the previous example. From Table 3.5, all models have comparable maximum log-likelihood values except MG. This may be due to the fact that the residual field seems smooth. Regarding prediction, C and WG have better performances than MC and WC, respectively as in Table 3.6. Note that WC has much larger sample standard deviations of estimates for support and shape parameters than WG.

When comparing the two mean structures, prediction errors for the constant mean

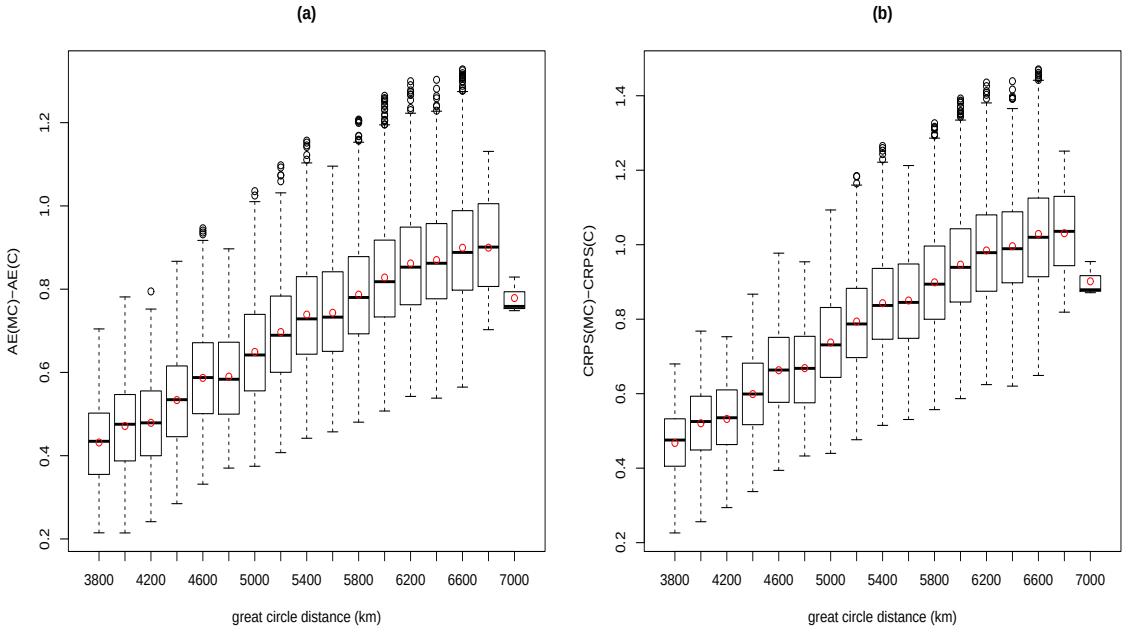

Figure 3.7: (Data example I - the constant mean) Boxplots of differences of AE (a) and CRPS (b) values from MC and C, displayed against minimum great circle distance between a prediction location and its nearest sampling location. Red circles represent average values in each bin.

Table 3.5: (Data example I - simple harmonic regression depending on latitude) Sample means and standard deviations of parameter estimates and maximum log-likelihood values for each model (100 cases). For models C and WG, $c \in (0, \pi]$ is a support parameter and $\tau \geq 6$ is a shape parameter. For model WG, $c > 0$ is a support parameter.

Model	$\hat{\sigma}^2$	$\hat{\alpha}$ or $\hat{\lambda}$	$\hat{\nu}$ or $\hat{c}$	$\hat{\tau}$	Max.loglik
MC	0.78(0.08)	1770.35(121.07)	2.70(0.09)	·	2171.23(21.33)
MG	0.10(0.00)	10754.17(91.31)	0.50(0.00)	·	1242.18(6.18)
C	0.82(0.05)	0.80(0.06)	2.89(0.07)	9.43(0.23)	2168.63(6.18)
WG	0.71(0.04)	·	2.87(0.02)	9.24(0.20)	2168.69(20.32)
WC	0.77(0.14)	·	9.79(22.49)	29.99(68.23)	2168.65(20.27)

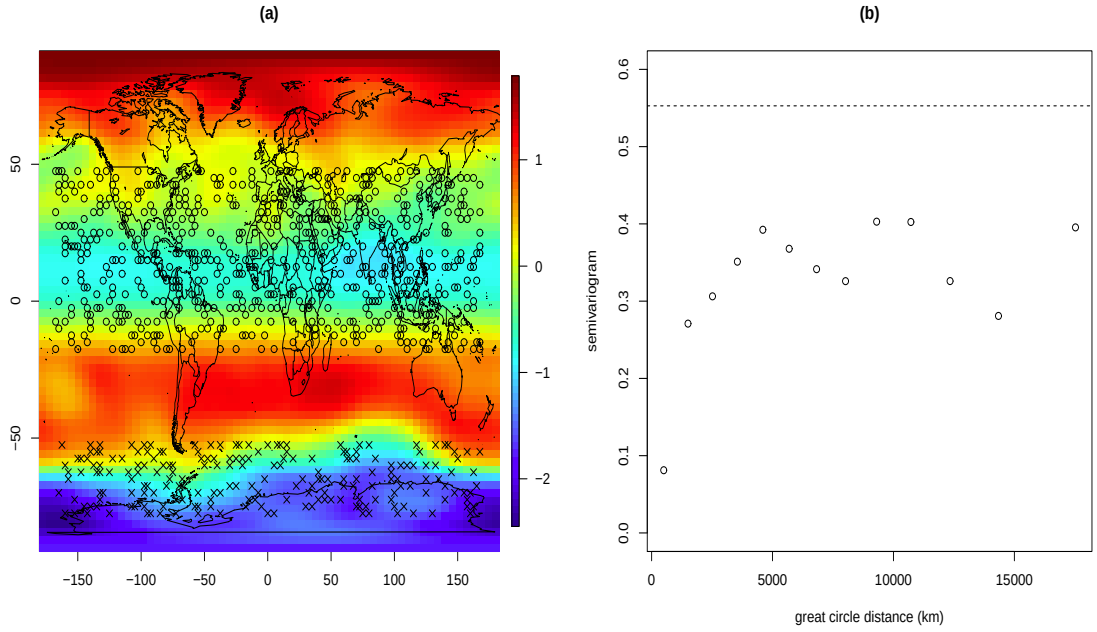

Figure 3.8: (Data example I) (a) A realization of residual fields after removing mean structure through simple harmonic regression depending on latitude. (a) shows sampling locations ($\circ$) and prediction locations ($\times$). (b) Empirical semivariogram values for selected locations displayed against the great circle distance. For (b), dotted line represents sample variance.

Table 3.6: (Data example I - simple harmonic regression depending on latitude) Sample means and standard deviations of prediction errors for each model (100 cases). For models C and WG, $c \in (0, \pi]$ is a support parameter and $\tau \geq 6$ is a shape parameter. For model WG, $c > 0$ is a support parameter.

Model	RMSPE	MAE	CRPS
MC	2.441(0.137)	2.314(0.142)	1.952(0.135)
MG	2.351(0.057)	2.217(0.065)	2.100(0.064)
C	2.177(0.085)	2.047(0.090)	1.677(0.084)
WG	1.988(0.074)	1.856(0.079)	1.493(0.072)
WC	2.184(0.357)	2.053(0.365)	1.689(0.351)

case are larger than those for the mean given by (3.8). Because the geopotential height data mainly show large scale, smooth, variation depending on latitude, the mean structure using simple harmonic regression resulted in improved prediction. Overall, the covariance functions of the class Ψ_d performed better than those of the class Φ_{d+1} regardless of mean structures.

3.4.3 Example II: vertical directional sampling design for prediction

We entertain the same set of covariance models with the mean structures as in Section 3.4.2. However, we changed a sampling design for prediction. We randomly selected 600 locations where longitude is less than 0° for parameter estimation, and selected 200 locations where that is greater than 0° for prediction as in Figure 3.9. The empirical semivariogram in Figure 3.9 shows that there is not much non-negative covariance values, unlike the previous example.

With a constant mean structure, Tables 3.7 and 3.8 show similar results as in Section 3.4.2. All models except MG are similar in terms of maximum log-likelihood values. Both C and WG have better performance in terms of prediction than MC and WC, respectively. Although results in Table 3.8 show that models C and WG do not outperform MC and WC as significantly as in Section 3.4.2, there still exist some improvements in terms of prediction with the models defined with the great circle distance. Figure 3.10 shows that WG has smaller AE and CRPS values than WC, when the prediction locations who are relatively far away from their nearest sampling locations.

When we consider the mean structure in (3.8), there is no significant difference in terms of prediction errors between C^4-Wendland models using the great circle distance and the Euclidean distance. However, both convex sum and C^4-Wendland models using the great circle distance perform better than Matérn covariance model

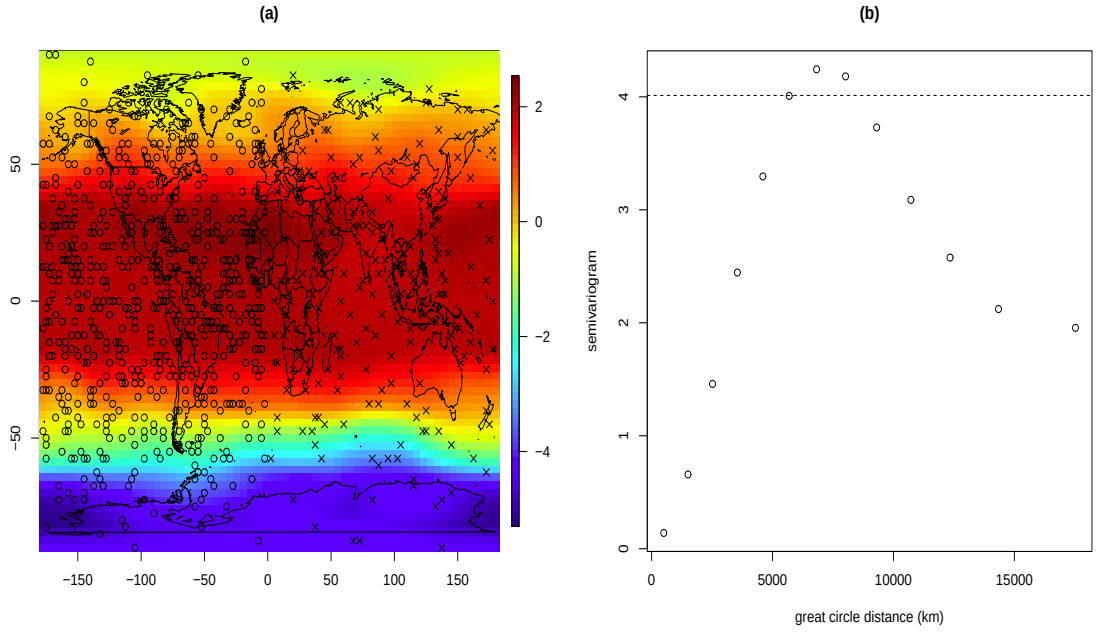

Figure 3.9: (Data example II) (a) A realization of residual fields after subtracting the constant mean. (a) shows sampling locations (∘) and prediction locations (×). (b) Empirical semivariogram values for selected locations displayed against the great circle distance. For (b), dotted line represents sample variance.

Table 3.7: (Data example II - constant mean) Sample means and standard deviations of parameter estimates, maximum log-likelihood values for each model (100 cases). For models C and WG, $c \in (0, \pi]$ is a support parameter and $\tau \geq 6$ is a shape parameter. For model WG, $c > 0$ is a support parameter.

Model	$\hat{\sigma}^2$	$\hat{\alpha}$ or $\hat{\lambda}$	$\hat{\nu}$ or $\hat{c}$	$\hat{\tau}$	Max.loglik
MC	2.06(0.28)	1220.49(158.69)	3.41(0.23)	·	1959.17(40.79)
MG	0.31(0.01)	10527.39(87.84)	0.50(0.00)	·	921.72(12.42)
C	3.62(0.32)	0.93(0.08)	2.94(0.17)	8.54(0.65)	1923.248(31.61)
WG	3.55(0.30)	·	3.03(0.06)	8.69(0.38)	1923.30(31.59)
WC	3.25(0.21)	·	3.02(0.96)	8.78(2.83)	1924.99(32.04)

Table 3.8: (Data example II - constant mean) Sample means and standard deviations of prediction errors for each model (100 cases). For models C and WG, $c \in (0, \pi]$ is a support parameter and $\tau \geq 6$ is a shape parameter. For model WG, $c > 0$ is a support parameter.

Model	RMSPE	MAE	CRPS
MC	1.167(0.079)	0.929(0.075)	0.606(0.054)
MG	1.610(0.063)	1.340(0.063)	1.167(0.059)
C	1.107(0.118)	0.855(0.105)	0.558(0.064)
WG	1.042(0.090)	0.804(0.086)	0.527(0.053)
WC	1.068(0.097)	0.832(0.090)	0.541(0.057)

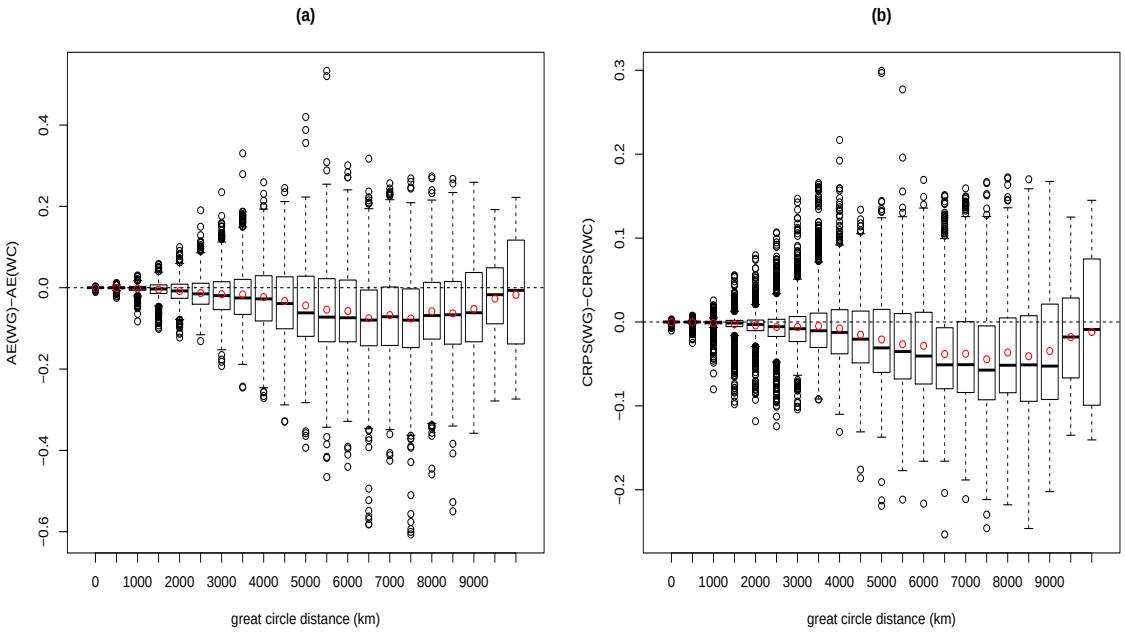

Figure 3.10: (Data example II - the constant mean) Boxplots of differences of AE (a) and CRPS (b) values from WG and WC, displayed against minimum great circle distance between a prediction location and its nearest sampling location. Red circles represent average values in each bin.

using the Euclidean distance in prediction. It is expected that the vertical directional design has smaller prediction errors than the horizontal directional design in Section 3.4.2 (Tables 3.4 and 3.8) because of large-scale variation depending strongly on latitude.

3.5 Conclusion

We have considered several classes of isotropic covariance functions with either the great circle distance or the Euclidean distance and compared them in terms of parameter estimation and spatial prediction. We have shown that when the true spatial range is large, the prediction performance of covariance models defined on the sphere using the great circle distance (that is, $\psi(\theta)$ on $\mathcal{S}_r^d$) is better than that of the functions projected from the Euclidean space. Moreover, when the data show significantly negative correlations at large distance lags, isotropic covariance models in the class Φ_3 are not adequate, and there is a substantial difference between covariance models from the classes Ψ_2 and Φ_3 in prediction. In the analysis of geopotential height data, we showed that distortion of the Euclidean distance might lead to poor prediction when the prediction locations are far away from sampling locations.

4. CROSS-COVARIANCE FUNCTIONS FOR SPACE-TIME PROCESSES ON A SPHERE

4.1 Introduction

The occurrence of multivariate data indexed by spatial coordinates and time points in various applications has prompted interest in statistical modeling in recent years. For examples, in climate and environmental sciences, climate variables such as temperature, pressure, and precipitation are collected by satellite monitors and they often vary in both space and time. Similarly, the output of climate models generates multiple variables. The primary interest in modeling such data is in detecting not only spatial-temporal dependence pattern but also relationship between distinct variables.

Covariance models for spatial-temporal data observed at locations on the surface of the Earth should be positive definite crossed with time. The covariance function for a space-time process that is space-time asymmetric is important for a better description of the process and prediction. The space-time process Z is called fully symmetric if $\text{cov}\{Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)\} = \text{cov}\{Z(\mathbf{s}_1, t_2), Z(\mathbf{s}_2, t_1)\}$ for all spatial locations $\mathbf{s}_1$ and $\mathbf{s}_2$ and time points t_1 and t_2 (Gneiting, 2002). If these covariances are not always the same, we call the space-time process Z space-time asymmetry. Jun and Stein (2007) proposed an approach to construct nonstationary covariance functions for a univariate process on the surface of a sphere, and demonstrated the effectiveness of their approach in capturing such space-time asymmetry.

Moreover, it has become popular to model multiple processes jointly. With multiple variables, prediction becomes a multiple problem and the multivariate extension of kriging (co-kriging) can be used. It is well known that co-kriging is often useful

when one variable is correlated with other variables that are readily observed. The Cross-covariance function is critical for the relationship between distinct variables and co-kriging. There have been noticeable advancements in developing parametric covariance models for multivariate spatial processes (Gneiting et al., 2010; Apanasovich and Genton, 2010; Apanasovich et al., 2012; Kleiber and Nychka, 2012; Jun, 2014). Genton and Kleiber (2015) would be of interest too.

In this Section, we extend multivariate Matérn models in Gneiting et al. (2010) for a space-time process on the surface of a sphere. We present a method to compute the approximate likelihood efficiently for the case of regularly spaced data of large dimension. We also illustrate our covariance models on a bivariate example of temperature and surface pressure.

4.2 Method

4.2.1 The isotropic Matérn cross-covariance model on $\mathcal{S}_R^2$

Gneiting et al. (2010) introduced the multivariate Matérn model to simultaneously model multiple processes, where each marginal process has a Matérn covariance structure. Suppose that a multivariate process $\mathbf{Z}(\mathbf{s}) = \{Z_1(\mathbf{s}), \ldots, Z_p(\mathbf{s})\}$ is observed over the Earth ($\mathbf{s} = (L, l) \in \mathcal{S}_R^2$, where L and l denote latitude and longitude, respectively, and R denotes the Earth's radius) and $K_{ij}(\mathbf{s}_1, \mathbf{s}_2) = \text{cov}\{Z_i(\mathbf{s}_1), Z_j(\mathbf{s}_2)\}$, then each marginal and cross-covariance functions are given by

$$\begin{aligned} K_{ii}(\mathbf{s}_1, \mathbf{s}_2) &= \sigma_i^2 \mathcal{M}_{\nu_i}\{d(\mathbf{s}_1, \mathbf{s}_2)/\beta_i)\},\ i = 1, \ldots, k, \\ K_{ij}(\mathbf{s}_1, \mathbf{s}_2) &= \rho_{ij}\sigma_i\sigma_j \mathcal{M}_{\nu_{ij}}\{d(\mathbf{s}_1, \mathbf{s}_2)/\beta_{ij})\},\ i \neq j, \end{aligned} \tag{4.1}$$

with $\mathcal{M}_{\nu_i}(x) = x^{\nu_i}\mathcal{K}_{\nu_i}(x)$ where $\mathcal{K}_\nu$ a modified Bessel function of the second kind with order ν_i, marginal variance $\sigma_i^2 > 0$, smoothness parameter $\nu_i > 0$, spatial

range parameter $\beta_i > 0$, and the collocated cross-correlation coefficient β_{ij} (it measures the strength of correlation between two processes at the same location). Here, $d(\mathbf{s}_1, \mathbf{s}_2) = 2R \times [\sin^2\{(L_1 - L_2)/2\} + \cos L_1 \cos L_2 \sin^2\{(l_1 - l_2)/2\}]^{0.5}$ is the chordal distance between two locations, $\mathbf{s}_1$ and $\mathbf{s}_2$. Gneiting et al. (2010) described two main multivariate Matérn models, the parsimonious and the full bivariate Matérn. In particular, the parsimonious Matérn has the same value of the spatial ranges, $\beta_i = \beta_{ij} = \beta$, for all marginal and cross-covariances, and the cross-smoothnesses are set to the $\nu_{ij} = (\nu_i + \nu_j)/2$. The model is valid under an easy-to-verify condition on ρ_{ij}. The assumption of common spatial range in this model is justified for some data sets (i.e. the example of temperature and pressure observations in Gneiting et al. (2010)). The full bivariate Matérn allows for different smoothness and spatial range parameters for two processes. For $k > 2$, the flexible Matérn is also introduced by Apanasovich et al. (2012).

4.2.2 The Matérn cross-covariance model on $\mathcal{S}_R^2 \times \mathbb{R}$

When adding the time dimension, the Matérn covariance function can be extended to the spatial-temporal setting. Suppose the a multivariate process $\mathbf{Z}(\mathbf{s}, t) = \{Z_1(\mathbf{s}, t), \ldots, Z_p(\mathbf{s}, t)\}$ is observed over the Earth for several time points ($(\mathbf{s}, t) = (L, l, t) \in \mathcal{S}_R^2 \times \mathbb{R}$). As in Jun and Stein (2007), we apply the restriction idea of Yadrenko (1983) and Yaglom (1987) (valid covariance functions on $\mathbb{R}^3$ can be restricted to $\mathcal{S}_R^2$ when the chordal distance is used) in order to get a valid spatial-temporal covariance functions on $\mathcal{S}_R^2 \times \mathbb{R}$. For spatial and temporal range parameters $\beta_s, \beta_t > 0$, we define the distance between two observations, $(\mathbf{s}_1, t_1)$ and $(\mathbf{s}_2, t_2)$, as given by

$$h(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2 | \beta_s, \beta_t) = \sqrt{d(\mathbf{s}_1, \mathbf{s}_2)^2/\beta_s^2 + |t_1 - t_2|^2/\beta_t^2}. \tag{4.2}$$

Then, $\sigma_i^2 \mathcal{M}_{\nu_i}(h)$ is a valid covariance function for a process on $\mathcal{S}_R^2 \times \mathbb{R}$. The parsimonious Matérn cross-covariance function can be extended to this setting straightforwardly. We assume that $\beta_{s,i} = \beta_{s,j} = \beta_{s,ij} = \beta_s$, $\beta_{t,i} = \beta_{t,j} = \beta_{t,ij} = \beta_t$ and $\nu_{ij} = (\nu_i + \nu_j)/2$. We may consider distinct process-dependent spatial or temporal range parameters for this model. Although it is not clear what conditions on these range parameters are required and further investigations are needed, the full bivariate Matérn model under this setting is at least valid when $\beta_{s,i}/\beta_{t,i} = \beta_{s,j}/\beta_{t,j} = \beta_{s,ij}/\beta_{t,ij}$ where $\beta_{s,ij} = \sqrt{(\beta_{s,i}^2 + \beta_{s,j}^2)/2}$ and $\beta_{t,ij} = \sqrt{(\beta_{t,i}^2 + \beta_{t,j}^2)/2}$. Here, we focus on spatially varying the marginal variance over space and our base covariance functions are

$$
\begin{aligned}
K_{ii}(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2) &= \sigma_i(\mathbf{s}_1)\sigma_i(\mathbf{s}_2)\mathcal{M}_{\nu_i}\{h(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2 | \beta_{s,i}, \beta_{t,i})\},\ i = 1, \ldots, k, \\
K_{ij}(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2) &= \rho_{ij}\sigma_i(\mathbf{s}_1)\sigma_j(\mathbf{s}_2)\mathcal{M}_{\nu_{ij}}\{h(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2 | \beta_{s,ij}, \beta_{t,ij})\},\ i \neq j, \qquad (4.3)
\end{aligned}
$$

The function $\sigma_i(\cdot)$ can be a positive function that varies over space. For example, we may consider the function has different values over the land and the sea or across latitude.

4.2.3 Computation

It is popular to use maximum likelihood estimation for the covariance parameters under the assumption that the process is multivariate Gaussian. A number of challenges arise when performing parameter estimation and prediction for large data sets because they typically require $\mathcal{O}(n^3)$ operations for a spatial-temporal data of size n. Inverting the covariance matrix and calculating the likelihood are often infeasible. However, for the regularly spaced data sets such as many satellite measurements of climate variables, the computation of the full likelihood can be efficient. Jun

and Stein (2008) and Jun (2011) demonstrate how to compute the exact likelihood through the Discrete Fourier Transform (DFT) for both univariate and bivariate processes of large dimension. The key idea is that for spatial data, when the covariance function is axially symmetric (the covariance structure is stationary with respect to longitude) and longitudinal values are equally spaced over the full longitude range (from -180° to 180°), the resulting covariance matrix can be written by a block circulant form. Then a block circulant matrix can be diagonalized by applying the DFT (Davis, 1979). Moreover, this matrix can be a block diagonal matrix easily with columns and rows rearranged. For example, suppose a spatial process is observed on a regular grid with q longitude points and p latitude points. Then the resulting covariance matrix after applying DFT becomes a $q \times q$ block diagonal matrix whose block size is $p \times p$, and the inverse and the determinant of the covariance matrix can be computed very efficiently (see Jun and Stein (2008) and Jun (2011) for more details).

When we add the time dimension, we can use a similar idea to compute the likelihood. It is well known that the Discrete Fourier Transform (DFT) has transformed a second order stationary time series into a sequence which is approximately uncorrelated (Dwivedi and Subba Rao, 2011) and thus the resulting covariance matrix is almost a diagonal. For a temporal process, which is observed on r time lags, the resulting covariance matrix by applying DFT becomes an almost diagonal matrix under the stationary assumption. If we ignore the relatively small off-diagonal elements (whose values approach 0 as $r \to \infty$), then we can compute the inverse and the determinant efficiently.

Now we combine these two ideas and compute the approximate likelihood for a space-time process. We consider the observed process $Z(\mathbf{s}, t)$ on a regular grid with q longitude points and p latitude points over r time points. Let us denote

$\mathbf{Z} = \{Z(t_i, l_j, L_k) : i = 1, \ldots, r, j = 1, \ldots, q, k = 1, \ldots, p\}^T$ (observations are ordered by time, then by longitude, and then by latitude) and the corresponding covariance matrix $V_{(r\times q\times p)\times(r\times q\times p)}$ (V is a symmetric toeplitz matrix with block circulant matrices). Then, $(F_r \otimes F_q \otimes I_p)V(F_r \otimes F_q \otimes I_p)^*$ becomes an almost $qr \times qr$ block diagonal matrix with $p \times p$ block diagonals where $F_N = (\omega^{jk}/\sqrt{N})_{j,k=0,\ldots,N-1}$ is the $N \times N$ DFT matrix, where $\omega = e^{-2\pi i/N}$ is a primitive Nth root of unity, and I_N is the $N \times N$ identity matrix. Relatively small off-block diagonal elements can be negligible for large r and in practice, the approximate likelihood computations through the DFT would require much smaller memory and less computation than the case of the full likelihood.

Suppose we consider a bivariate space-time process $\{Z_1(\mathbf{s}, t), Z_2(\mathbf{s}, t)\}$ observed on a regular grid with q longitude points (covering full range), p latitude points, and r time points. If we denote $\mathbf{Z}^* = \{\mathbf{Z}_1^T, \mathbf{Z}_2^T\}^T$ where $\mathbf{Z}_l = \{Z_l(t_i, l_j, L_k) : i = 1, \ldots, r, j = 1, \ldots, q, k = 1, \ldots, p\}^T$ for $l = 1, 2$, then the covariance matrix of $\mathbf{Z}^*$ can be written as

$$\Sigma = \begin{bmatrix} V_1 & V_{12} \\ V_{12}^* & V_2 \end{bmatrix}$$

where V_1, V_2, and V_{12} are $qr \times qr$ complex block diagonal matrices with $p \times p$ block diagonals and V_{12}^* is the conjugate transpose of $V12$. The determinant of Σ can be computed using $|\Sigma| = |V_1 - V_{12}V_2^{-1}V_{12}^*| \times |V_2|$ where $|V|$ is the determinant of the matrix V and the inverse of Σ can be computed efficiently

$$\Sigma^{-1} = \begin{bmatrix} (V_1 - V_{12}V_2^{-1}V_{12}^*)^{-1} & -V_1^{-1}V_{12}(V_2 - V_{12}^*V_2^{-1}V_{12})^{-1} \\ -V_2^{-1}V_{12}^*(V_1 - V_{12}V_2^{-1}V_{12}^*)^{-1} & (V_2 - V_{12}^*V_2^{-1}V_{12})^{-1} \end{bmatrix}.$$

For simulation and application, we will use these computational techniques under the

(second order) stationary assumption in time and estimate the covariance parameters based on the approximate likelihood.

4.3 Simulation study

We consider a Gaussian random field on $\mathcal{S}_R^2 \times \mathbb{R}$ with mean 0 and a bivariate parsimonious Matérn covariance function with chordal distance in (4.3), but we set the variances are constant. We suppose a space-time process observed on a regular grid with $q = 24$ longitude points, $p = 13$ latitude points, and $r = 12$ time points over the surface of the Earth with $R = 6371$ (km). We set $\sigma_1^2 = 20, \sigma_2^2 = 30, \sigma_{12} = \sigma_1\sigma_2, \beta_s = 800$ (km), $\beta_t = 2, \nu_1 = 1.5, \nu_2 = 0.5, \nu_{12} = (\nu_1 + \nu_2)/2$, and $\rho_{12} = 0.3$. To guarantee the valid bivariate covariance function, we need a restriction for the cross-correlation parameter, $|\rho_{12}| \leq \Gamma(\frac{\nu_1+\nu_2}{2})\Gamma(\nu_1)^{-0.5}\Gamma(\nu_2)^{-0.5}$.

We used the maximum likelihood estimation method and compared fitted results from the approximate likelihood approach based on the DFT method with the assumption of stationarity in time (STA) described in Section 4.2.3 to those from the full likelihood approach (FLM) based on the bivariate parsimonious Matérn model. Moreover, we consider partitioning over time (IND): our observations are assumed to be independent across partitioned space-time subregions. Let $B_1, \ldots, B_l$ be a partition of the space-time domain $\mathcal{S}_R^2 \times [1, T]$ where $T = 12$, referred to as blocks. For IND, we consider $l = 3$ blocks with the same size.

Table 4.1 contains the model parameter estimates under each likelihood approach from the bivariate parsimonious Matérn model. For both approximate approaches, parameters are estimated well compared to the full likelihood approach. For IND, the smoothness parameter of ν_2 is overestimated compared to other approaches. For STA, the spatial range and the temporal range parameters are slightly underestimated. A potential explanation for the underestimated $\hat{\beta}_t$ here is that in the

Table 4.1: Parameter estimates for each method from the bivariate parsimonious Matérn model. $k = \beta_s/\beta_t = 400$ and $\nu_{12} = (\nu_1 + \nu_2)/2 = 1$. Computation time in seconds for calculating the log-likelihood value is also given.

Model	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\beta}_s$	$\hat{\beta}_t$	$\hat{\nu}_1$	$\hat{\nu}_2$	$\hat{\rho}_{12}$	TIME (s)
TRUE	20	30	800	2	1.5	0.5	0.3	·
STA	19.999	29.998	761.057	1.850	1.499	0.478	0.329	1.08
IND	19.999	29.997	801.588	2.011	1.485	0.635	0.339	0.69
FLM	20.343	30.691	843.553	2.023	1.478	0.504	0.324	136.07

current study, we assume that the covariance matrix is a block diagonal matrix and ignore some off-block diagonal elements of the covariance matrix under the stationary assumption in time and the DFT. This may mislead some information of the temporal dependencies. The off-block diagonal elements are negligible after the DFT when the number of time points is large enough, but here $r = 12$ may not large enough. Moreover, our covariance model depends on the distance, $h(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2|\beta_s, \beta_t) = \sqrt{d(\mathbf{s}_1, \mathbf{s}_2)^2/\beta_s^2 + |t_1 - t_2|^2/\beta_t^2} = \sqrt{d(\mathbf{s}_1, \mathbf{s}_2)^2 + k^2|t_1 - t_2|^2}/\beta_s$ where $k = \beta_s/\beta_t > 0$, and it seems that, for fixed number of time points, the approximation becomes accurate as the k increases. Since the magnitude of $d(\mathbf{s}_1, \mathbf{s}_2)$ is larger than that of $|t_1 - t_2|$ here, the distance $h(\cdot)$ is dominated by the spatial distance unless k is large enough and our method depends on spatial range mainly. When we set β_t is large compared to the number of time points, the temporal range parameters are strongly underestimated for both approximate approaches. The relationship between the magnitudes of spatial and temporal distances and the ratio of spatial and temporal range, $k = \beta_s/\beta_t$, need to be further investigated. Note that computation times of the two approximate likelihood approaches are much faster than the case of the full likelihood approach as in Table 4.1. The computation times were recorded on a Intel i5-2430M 2.40GHz CPU. Table 4.2 represents the parameter

estimates from the bivariate full Matérn model. The smoothness parameters for the two variables and the collocated correlation parameter are overestimated, but the estimates of the spatial range and the temporal range parameters are close to the true values unlike the example of the bivariate parsimonious Matérn model.

Table 4.2: Parameter estimates for each method from the bivariate full Matérn model. $k = \beta_{s1}/\beta_{t1} = \beta_{s2}/\beta_{t2} = \beta_{s12}/\beta_{t12} = 400$, $\beta_{s12} = \sqrt{(\beta_{s1}^2 + \beta_{s2}^2)/2}$, and $\nu_{12} = (\nu_1 + \nu_2)/2 = 1$.

Model	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\beta}_{s1}$	$\hat{\beta}_{s2}$	$\hat{\beta}_{t1}$	$\hat{\beta}_{t2}$	$\hat{\nu}_1$	$\hat{\nu}_2$	$\hat{\rho}_{12}$
TRUE	30	20	800	600	2	1.5	1.5	0.5	0.3
STA	20.008	29.976	826.548	602.541	2.160	1.574	1.760	0.549	0.348
IND	20.005	29.995	819.316	603.603	2.103	1.550	1.734	0.550	0.348
FLM	21.467	30.104	827.158	635.905	2.043	1.570	1.497	0.505	0.309

4.4 Application

Gneiting et al. (2010) applied their Matérn cross-covariance model to the temperature and pressure observations in the North American Pacific Northwest. It is well known that the observation fields are rough for temperature and smooth for pressure. Furthermore, the height of a given pressure surface, which is a surface in the atmosphere where the pressure remains the same everywhere along that surface, above the ground varies with temperature. Because a given pressure surface in a colder air will have a lower height than when the same pressure surface was located in warmer air, the temperature and surface pressure variables are negatively correlated significantly. Gneiting et al. (2010) showed that due to the clear dependence between the two variables (the collocated empirical correlation coefficient for the dataset is -0.47), their multivariate Matérn models outperform the independent

Matérn model.

We consider temperature and surface pressure data on a global scale. Data sets are obtained from the ERA-Interim reanalysis project (http://www.ecmwf.int/en/research/climate-reanalysis/era-interim), which is the latest global atmospheric reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). The data assimilation system used to produce ERA-Interim is based on a December 2006 version of the ECMWF Integrated Forecast Model (IFS-Cy31r2) (see Dee et al. (2011) for more details). The output values for each variable from June 1 to June 16, 2014 (daily outputs that are measured at 0 am) are given on regular grids ($3^\circ \times 3^\circ$), and there are 120 longitude points, 59 latitude points, and 16 time points (the size of the data is 113,280). The units are K for the temperature and kPa for the surface pressure.

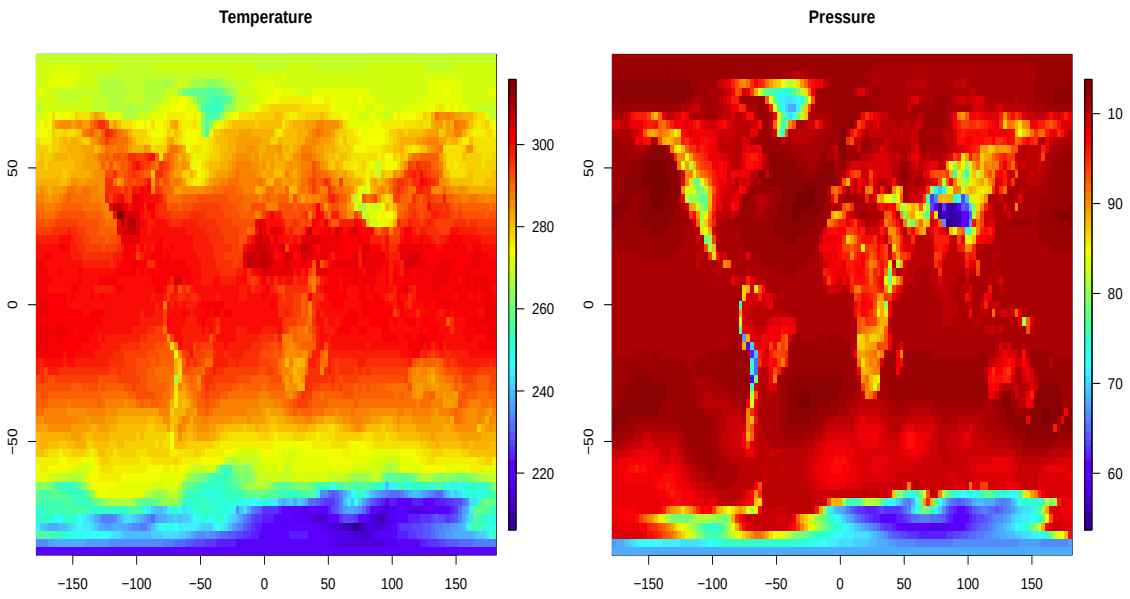

Figure 4.1: Temperature (K) and surface pressure (kPa) on June 1, 2014 at 12 am.

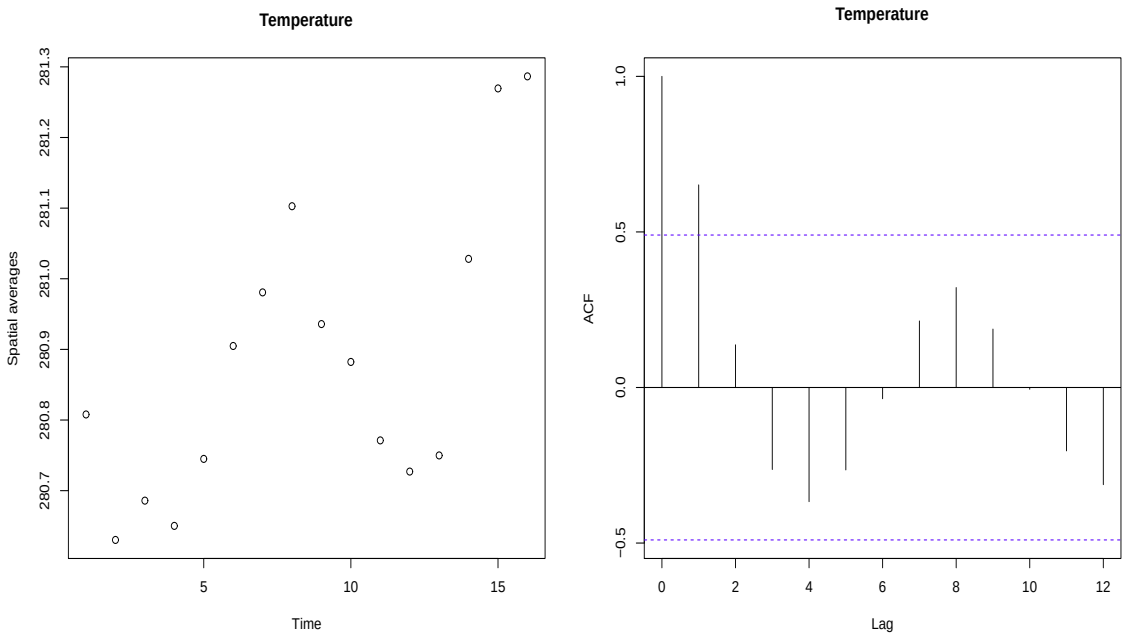

Figure 4.2: The spatial averages of the original temperature data against time and the corresponding sample ACFs.

We decompose the data into its mean structure and the residual. In Figure 4.1, the temperature variable suggests clear large scale spatial structure depending on latitude and the surface pressure variable has a strong dependence on altitude (i.e. Himalayas and Chile's mountain chains). Thus, for the spatial mean structure, we consider linear regression with covariates, altitude and quadratic polynomials of latitude. In Figures 4.2 and 4.3 we observe that the spatial averages (over all locations) displayed against time points and the corresponding temporal sample autocorrelation (ACF) plots for temperature and surface pressure. In Section 4.2.3 we have assumed that temporal (second order) stationarity to use the DFT method. But from the mean and ACF plots of Figures 4.2 and 4.3 it is clear that there are some deterministic trends. To remove the trends, we modeled the temporal mean structure

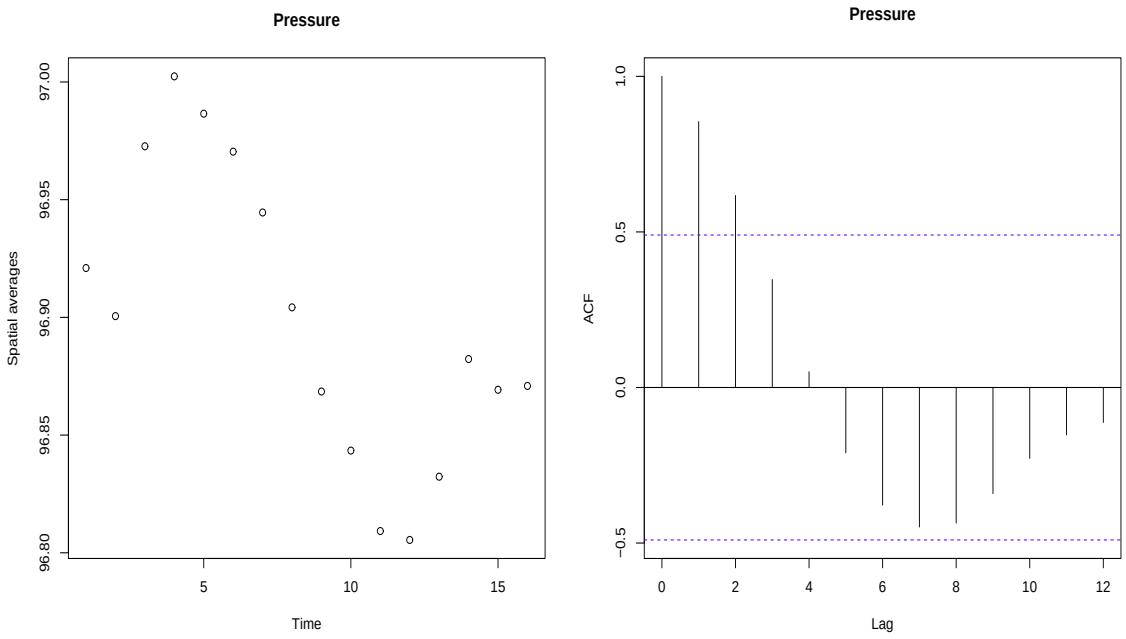

Figure 4.3: The spatial averages of the original surface pressure data against time and the corresponding sample ACFs.

through simple harmonic regression depending on time points. From residual fields of the two variables after subtracting the spatial and temporal mean structures in Figure 4.4, we observe that variations over the land are more noticeable than those over the sea.

We consider the parsimonious bivariate Matérn model with spatially varying variance. Since both temperature and pressure variables exhibit dependence of their covariance structure on latitude and land/sea factor, we allow variances to be different not only over the land and the sea but also across latitude. We set

$$
\begin{aligned}
Z_1(\mathbf{s},t) &= \{a_0 + a_1 Lat(\mathbf{s}) + a_2 1_{(\mathbf{s}\in \text{land})}\} \times Z_1^{\circ}(\mathbf{s},t), \\
Z_2(\mathbf{s},t) &= \{b_0 + b_1 Lat(\mathbf{s}) + b_2 1_{(\mathbf{s}\in \text{land})}\} \times Z_2^{\circ}(\mathbf{s},t),
\end{aligned} \tag{4.4}
$$

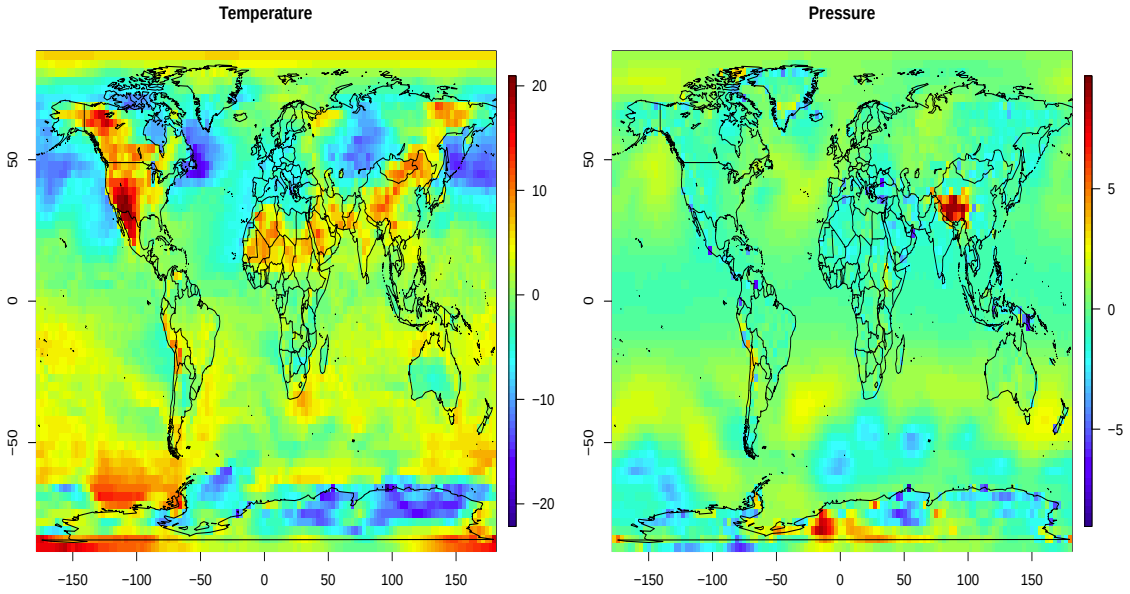

Figure 4.4: Residual fields of temperature and surface pressure on June 1, 2014 at 12 am.

where $Lat(\mathbf{s})$ is a function depending on the latitude, and we used finite linear combinations of Legendre polynomials as in Jun (2014), $Lat(\mathbf{s}) = \sum_{j=0}^{m} c_{ij} P_j(\sin L) \times \cos(L)$, $i = 1, 2$ and $m = 2$. To address an identifiability problem, we fixed $c_{i0} = -1$. We assume that the processes Z_i°, $i = 1, 2$, defined on $\mathcal{S}_R^2 \times \mathbb{R}$ has mean zero and its covariance is given by the parsimonious Matérn covariance function, $\mathcal{M}_{\nu_i}\{h(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2 | \beta_s, \beta_t)\}$. We also let the cross covariance between Z_i° and Z_j° is $\rho_{ij}\mathcal{M}_{\nu_{ij}}\{h(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2 | \beta_s, \beta_t)\}$.

The estimated covariance parameter values from the approximate approaches, STA and IND, are given in Table 4.3. It seems that the pressure process is smoother than the temperature process, which agrees with findings in Gneiting et al. (2010). The temporal range is only 1.46 days and the spatial range is about 370 km from STA.

Table 4.3: Parameter estimates for each method from the bivariate parsimonious Matérn model. For identifiability, we set $c_{10} = c_{20} = -1$.

Model	$\hat{a}_0$	$\hat{b}_0$	$\hat{a}_1$	$\hat{b}_1$	$\hat{a}_2$	$\hat{b}_2$	$\hat{\beta}_s$	$\hat{\beta}_t$
STA	3.924	0.988	3.904	0.986	2.768	0.698	372.715	1.460
IND	4.029	0.989	3.837	0.983	2.722	0.696	308.911	0.970
Model	$\hat{\nu}_1$	$\hat{\nu}_2$	$\hat{\rho}_{12}$	$\hat{c}_{11}$	$\hat{c}_{12}$	$\hat{c}_{21}$	$\hat{c}_{22}$	
STA	0.459	1.131	-0.213	-0.377	-0.375	0.118	0.016	
IND	0.451	1.123	-0.212	-2.014	-2.514	-0.418	-0.768	

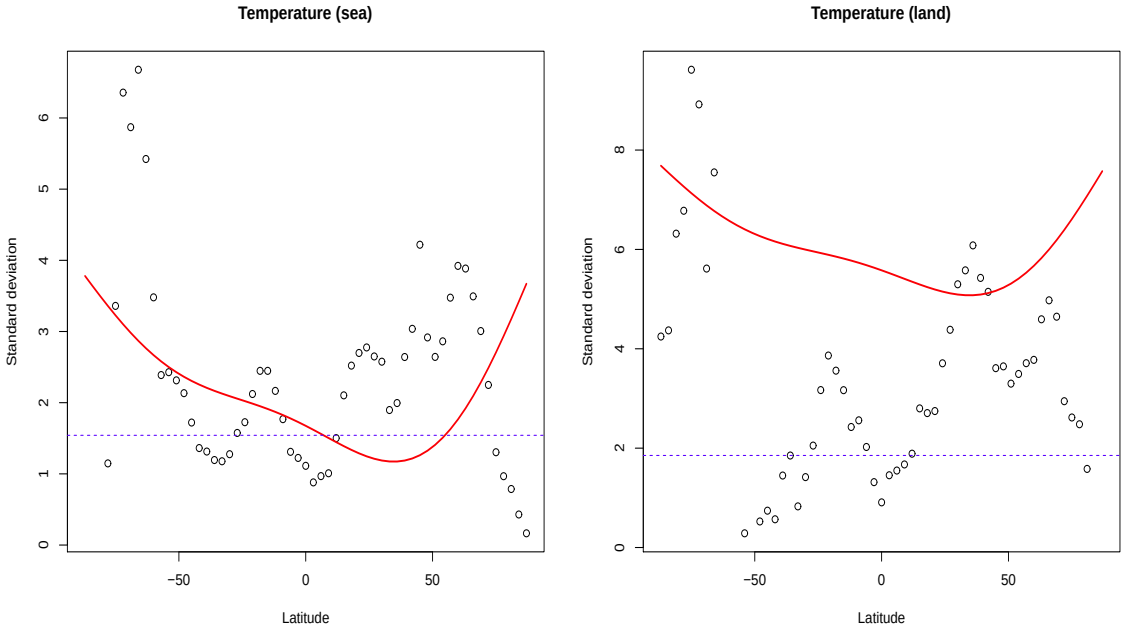

Figure 4.5: Comparison of empirical standard deviations for the temporal average of the temperature residuals over the sea and the land against latitude (dots) and corresponding fitted values (red lines). Fitted values are calculated using the covariance parameter estimates from STA. Blue dots represents empirical standard deviations over the sea and the land.

Figures 4.5 and 4.6 provide comparisons between the empirical and fitted standard deviations for the temporal average of the the temperature and the surface pressure

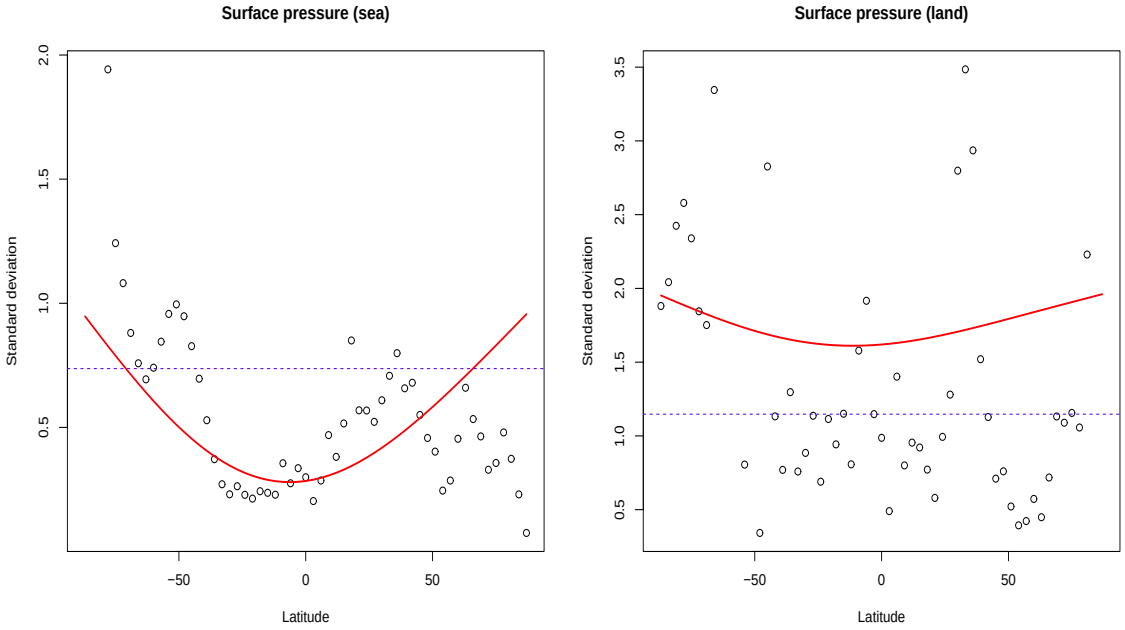

Figure 4.6: Comparison of empirical standard deviations for the temporal average of the surface pressure residuals over the sea and the land against latitude (dots) and corresponding fitted values (red lines). Fitted values are calculated using the covariance parameter estimates from STA. Blue dots represents empirical standard deviations over the sea and the land.

residuals over the sea and the land against latitude. Fitted values work reasonably well to capture the pattern of the sample standard deviations over the sea and the land in the surface pressure data except arctic and antarctic areas. For a better fit, it is interesting to increase m in $Lat(\mathbf{s})$ for the temperature and the surface pressure data and the full bivariate model in (4.3) can be considered.

4.5 Conclusion

We presented the bivariate Matérn-based covariance model for space-time processes on the surface of a sphere. For spatial-temporal data, we showed how the DFT approach can be used to save memory and reduce computational burden. Al-

though we have focused on the spatially varying marginal variance parameters in the bivariate Maérn model, we fixed the constant smoothness parameters over space and it may be restrictive for global climate variables. As Jun (2014) demonstrated the effectiveness of spatially varying smoothness parameters in the example of surface temperature and precipitation, allowing smoothness parameter to be different over space would improve the fit of the models remarkably.

An interesting direction for the future work is to extend multivariate covariance models in Jun (2011, 2014) for a space-time process on the surface of a sphere and demonstrate their application to the fine scale spatial-temporal data. They showed how the nonstationary Matérn model can be coupled with the differential operators approach. Our focus should be to consider the covariance model which is flexible enough to capture not only space-time asymmetry but also nonstationarity in the data.

5. SUMMARY

In this dissertation, we explore spatial-temporal covariance models on the surface of a sphere in climate problem. This topic is an important issue in the analysis of complex physical processes on the globe such as satellite measurements of climate variables. The theoretical challenge in modeling spatial processes lies in formulating a valid covariance function. This motivates the study of random fields and valid covariance functions on the surface of a sphere.

In Section 2, we introduce a statistical methodology to develop parametric covariance functions using the great circle distance for spatial processes, geared towards smooth processes on the surface of a sphere. The key idea is to integrate a non-differential process over a small neighborhood on the surface of a sphere, which results in a smoother process. The resulting model is isotropic and positive definite on the surface of a sphere. The method proposed offers a natural nonstationary extension for more flexible covariance structure.

In Section 3, we consider a comparative study to compare isotropic covariance functions using the either the great circle distance or the Euclidean distance on the surface of a sphere. Several classes of covariance functions on the surface of a sphere, defined with either the great circle distance or the Euclidean distance, are considered and their impact upon prediction is investigated. We demonstrate that covariance functions originally defined by the Euclidean distance may be limited to global data.

In Section 4, we consider cross-covariance models for spatial-temporal data on the surface of a sphere. In order to guarantee the positive definiteness and flexibility of the covariance function, we extend the bivariate Matérn models using the chordal distance for space-time processes and consider the spatially varying marginal variance

parameters in the models. We illustrate a method to compute the approximate likelihood efficiently for the case of regularly spaced data of large dimension.

REFERENCES

Apanasovich, T. V. and Genton, M. G. (2010). Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika* **97,** 15–30.

Apanasovich, T. V., Genton, M. G., and Sun, Y. (2012). A valid matérn class of cross-covariance functions for multivariate random fields with any number of components. *Journal of the American Statistical Association* **107,** 180–193.

Banerjee, S. (2005). On geodetic distance computations in spatial modeling. *Biometrics* **61,** 617–625.

Chen, D., Menegatto, V. A., and Sun, X. (2003). A necessary and sufficient condition for strictly positive definite functions on spheres. *Proceedings of the American Mathematical Society* **13,** 2733–2740.

Davis, P. J. (1979). *Circulant matrices*. John Wiley & Sons, New York.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al. (2011). The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* **137,** 553–597.

Du, J. and Ma, C. (2012). Variogram matrix functions for vector random fields with second-order increments. *Mathematical Geosciences* **44,** 411–425.

Du, J., Ma, C., and Li, Y. (2013). Isotropic variogram matrix functions on spheres. *Mathematical Geosciences* **45,** 341–357.

Dwivedi, Y. and Subba Rao, S. (2011). A test for second-order stationarity of a time series based on the discrete fourier transform. *Journal of Time Series Analysis* **32,** 68–91.

Furrer, R., Genton, M. G., and Nychka, D. (2006). Covariance tapering for interpo-

lation of large spatial datasets. *Journal of Computational and Graphical Statistics* **15,** 502–523.

Genton, M. and Kleiber, W. (2015). Cross-covariance functions for multivariate geostatistics (with discussion). *Statistical Science* **30,** 147–163.

Gneiting, T. (1999). Correlation functions for atmospheric data analysis. *Quarterly Journal of the Royal Meteorological Society* **125,** 2449–2464.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association* **97,** 590–600.

Gneiting, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli* **19,** 1327–1349.

Gneiting, T., Kleiber, W., and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association* **105,** 180–193.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **102,** 359–378.

Guinness, J. and Fuentes, M. (2013). Covariance functions for mean square differentiable processes on spheres. Technical report, Department of Statistics, NC State Univ. Available at http://www.stat.ncsu.edu/information/library/mimeo.php.

Hafez, Y. (2012). In Yucel, I., editor, *Atmospheric Model Application*, chapter Blocking Systems Persist over North Hemisphere and Its Role in Extreme Hot Waves over Russia during Summer 2010. INTECH Open Access Publisher, Rijeka.

Hafez, Y. Y. and Almazroui, M. (2014). Recent study of anomaly of global annual geopotential height and global warming. *Atmospheric and Climate Sciences* **4,** 347–357.

Hansen, L. V., Thorarinsdottir, T. L., and Gneiting, T. (2011). Lévy particles: Modelling and simulating star-shaped random sets. Technical report, Research Report

2011/04, Centre for Stochastic Geometry and Advanced Bioimaging, Aarhus Univ. Available at http://data.imf.au.dk/publications/csgb/2011/imf-csgb-2011-04. pdf.

Heaton, M., Katzfuss, M., Berrett, C., and Nychka, D. (2014). Constructing valid spatial processes on the sphere using kernel convolutions. *Environmetrics* **25,** 2–15.

Huang, C., Zhang, H., and Robeson, S. M. (2011). On the validity of commonly used covariance and variogram functions on the sphere. *Mathematical Geosciences* **43,** 721–733.

Jeong, J. and Jun, M. (2015). A class of matérn-like covariance functions for smooth processes on a sphere. *Spatial Statistics* **11,** 1–18.

Jun, M. (2011). Non-stationary cross-covariance models for multivariate processes on a globe. *Scandinavian Journal of Statistics* **38,** 726–747.

Jun, M. (2014). Matérn-based nonstationary cross-covariance models for global processes. *Journal of Multivariate Analysis* **128,** 134–146.

Jun, M. and Stein, M. L. (2007). An approach to producing space–time covariance functions on spheres. *Technometrics* **49,** 468–479.

Jun, M. and Stein, M. L. (2008). Nonstationary covariance models for global data. *The Annals of Applied Statistics* **2,** 1271–1289.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., et al. (1996). The ncep/ncar 40-year reanalysis project. *Bulletin of the American Meteorological Society* **77,** 437–471.

Kaufman, C. G., Schervish, M. J., and Nychka, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association* **103,** 1545–1555.

Kleiber, W. and Nychka, D. (2012). Nonstationary modeling for multivariate spatial processes. *Journal of Multivariate Analysis* **112,** 76–91.

Marshall, G. J. (2002). Trends in antarctic geopotential height and temperature: A

comparison between radiosonde and ncep-ncar reanalysis data. *Journal of Climate* **15,** 659–674.

Martins, T. G., Simpson, D., Lindgren, F., and Rue, H. (2013). Bayesian computing with inla: new features. *Computational Statistics & Data Analysis* **67,** 68–83.

Miller, K. S. and Samko, S. G. (2001). Completely monotonic functions. *Integral Transforms and Special Functions* **12,** 389–402.

Paciorek, C. J. and Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* **17,** 483–506.

Parzen, E. (1999). *Stochastic processes*. Society for Industrial and Applied Mathematics, Philadelphia.

Sang, H. and Huang, J. Z. (2012). A full scale approximation of covariance functions for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74,** 111–132.

Schoenberg, I. (1942). Positive definite functions on spheres. *Duke Mathematical Journal* **9,** 96–108.

Stein, M. L. (1999). *Interpolation of spatial data: some theory for kriging*. Springer-Verlag, New York.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012). An overview of cmip5 and the experiment design. *Bulletin of the American Meteorological Society* **93,** 485–498.

Wendland, H. (1995). Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in computational Mathematics* **4,** 389–396.

Yadrenko, M. I. (1983). *Spectral theory of random fields*. Optimization Software, New York.

Yaglom, A. (1987). *Correlation Theory of Stationary and Related Random Functions: Vol.: 1: Basic Results*. Springer-Verlag, New York.

Zhu, J., Wang, S., and Zhang, X. (2002). Global warming mode of atmospheric circulation. *Atmospheric Science Letters* **3,** 1–13.

APPENDIX A

SUPPLEMENT TO THE DERIVATION OF EQUATION (2.5) IN SECTION 2

We consider $W(s) = \int_{s-\epsilon}^{s+\epsilon} Z(u)du/(2\epsilon)$ on $\mathcal{S}^1$ with zero mean and $C(h) = e^{-h/\alpha}$ so that $\mathrm{cov}\{W(s), W(t)\} = \int_{t-\epsilon}^{t+\epsilon}\int_{s-\epsilon}^{s+\epsilon} e^{-|u-v|/\alpha}dudv \Big/ (2\epsilon)^2$.

For $|t-s| \le \epsilon$ or $\epsilon \le |t-s| \le 2\epsilon$ $(s < t)$,

$$\int_{t-\epsilon}^{s+\epsilon}\{\int_{s-\epsilon}^{v} e^{-(v-u)/\alpha}du + \int_{v}^{s+\epsilon} e^{-(u-v)/\alpha}du\}dv + \int_{s+\epsilon}^{t+\epsilon}\int_{s-\epsilon}^{s+\epsilon} e^{-(v-u)/\alpha}dudv$$

$$= \int_{t-\epsilon}^{s+\epsilon} \alpha e^{-v/\alpha}\{e^{v/\alpha} - e^{(s-\epsilon)/\alpha}\} + \alpha e^{v/\alpha}\{e^{-v/\alpha} - e^{-(s+\epsilon)/\alpha}\}dv$$

$$+ \int_{s+\epsilon}^{t+\epsilon} \alpha e^{-v/\alpha}\{e^{(s+\epsilon)/\alpha} - e^{(s-\epsilon)/\alpha}\}dv$$

$$= \alpha\{v + e^{\frac{-v+(s-\epsilon)}{\alpha}}\}\Big|_{t-\epsilon}^{s+\epsilon} + \alpha\{v - \alpha e^{\frac{v-(s+\epsilon)}{\alpha}}\}\Big|_{t-\epsilon}^{s+\epsilon}$$

$$+\alpha^2\{e^{(s+\epsilon)/\alpha} - e^{(s-\epsilon)/\alpha}\}\{e^{-(s+\epsilon)/\alpha} - e^{-(t+\epsilon)/\alpha}\}$$

$$= \alpha\{-(t-s) + 2\epsilon + \alpha e^{-2\epsilon/\alpha} - \alpha e^{-(t-s)/\alpha}\} + \alpha\{-(t-s) + 2\epsilon - \alpha + \alpha e^{\frac{(t-s)-2\epsilon}{\alpha}}\}$$

$$+\alpha^2\{1 - e^{-(t-s)/\alpha} - e^{-2\epsilon/\alpha} + e^{\frac{-(t-s)-2\epsilon}{\alpha}}\}$$

$$= \alpha\{4\epsilon - 2(t-s)\} + \alpha^2\{-2e^{-(t-s)/\alpha} + e^{\frac{(t-s)-2\epsilon}{\alpha}} + e^{\frac{-(t-s)-2\epsilon}{\alpha}}\}.$$

For $2\epsilon < |t-s| < \pi - \epsilon$ $(s < t)$,

$$\int_{t-\epsilon}^{t+\epsilon}\int_{s-\epsilon}^{s+\epsilon} e^{-(v-u)/\alpha}dudv = \int_{t-\epsilon}^{t+\epsilon} \alpha e^{-v/\alpha}\{e^{(s+\epsilon)/\alpha} - e^{(s-\epsilon)/\alpha}\}dv$$

$$= -\alpha^2\{e^{(s+\epsilon)/\alpha} - e^{(s-\epsilon)/\alpha}\}e^{-v/\alpha}\Big|_{t-\epsilon}^{t+\epsilon}$$

$$= \alpha^2\{e^{(s+\epsilon)/\alpha} - e^{(s-\epsilon)/\alpha}\}\{e^{-(t-\epsilon)/\alpha} - e^{-(t+\epsilon)/\alpha}\} = \alpha^2 e^{-(t-s)/\alpha}\{e^{2\epsilon/\alpha} + e^{-2\epsilon/\alpha} - 2\}.$$

For $|t-s| \geq \pi - \epsilon$ $(s < t)$,

$$\int_{t-\epsilon}^{s+\pi}\int_{s-\epsilon}^{s+\epsilon} e^{-\frac{(v-u)}{\alpha}}dudv + \int_{s+\pi}^{t+\epsilon}\int_{s-\epsilon}^{s+\epsilon} e^{-\frac{2\pi-(v-u)}{\alpha}}dudv$$

$$= \int_{t-\epsilon}^{s+\pi} e^{-v/\alpha}\alpha\{e^{\frac{s+\epsilon}{\alpha}} - e^{\frac{s-\epsilon}{\alpha}}\} + \int_{s+\pi}^{t+\epsilon} e^{\frac{-2\pi+v}{\alpha}}\alpha\{e^{\frac{-s+\epsilon}{\alpha}} - e^{\frac{-s-\epsilon}{\alpha}}\}$$

$$= \alpha^2(e^{\frac{s+\epsilon}{\alpha}} - e^{\frac{s-\epsilon}{\alpha}})(e^{\frac{-t+\epsilon}{\alpha}} - e^{\frac{-s-\pi}{\alpha}}) + \alpha^2 e^{-2\pi/\alpha}(e^{\frac{-s+\epsilon}{\alpha}} - e^{\frac{-s-\epsilon}{\alpha}})(e^{\frac{t+\epsilon}{\alpha}} - e^{\frac{s+\pi}{\alpha}})$$

$$= \alpha^2\{-2e^{\frac{\epsilon-\pi}{\alpha}} + 2e^{\frac{-\pi-\epsilon}{\alpha}} + e^{\frac{-(t-s)}{\alpha}}(e^{2\epsilon/\alpha} - 1) + e^{\frac{(t-s)-2\pi}{\alpha}}(e^{2\epsilon/\alpha} - 1)\}.$$

From the aboves we have (2.5) with $\theta = |t-s|$,

$$(2\epsilon)^2\mathrm{cov}\{W(s), W(t)\}$$
$$= \begin{cases} \alpha(4\epsilon - 2\theta) + \alpha^2(-2e^{-\theta/\alpha} + e^{\frac{\theta-2\epsilon}{\alpha}} + e^{\frac{-\theta-2\epsilon}{\alpha}}) & \text{if } \theta \leq 2\epsilon \\ \alpha^2 e^{-\theta/\alpha}(e^{2\epsilon/\alpha} + e^{-2\epsilon/\alpha} - 2) & \text{if } 2\epsilon < \theta < \pi - \epsilon \\ \alpha^2\{-2e^{\frac{\epsilon-\pi}{\alpha}} + 2e^{\frac{-\pi-\epsilon}{\alpha}} + e^{\frac{-\theta}{\alpha}}(e^{2\epsilon/\alpha} - 1) + e^{\frac{\theta-2\pi}{\alpha}}(e^{2\epsilon/\alpha} - 1)\} & \text{if } \theta \geq \pi - \epsilon. \end{cases} \tag{A.1}$$

APPENDIX B

SUPPLEMENT TO THE DERIVATION OF EQUATION (2.6) IN SECTION 2

We want to derive the equation (2.5). We consider $K(h) = e^{-\theta/\alpha}$ where $\theta = |t-s|$ and $\text{cov}\{W(s), W(t)\} = \int_{t-\epsilon}^{t+\epsilon} \int_{s-\epsilon}^{s+\epsilon} K(|u-v|dudv/(2\epsilon)^2$. From (A.1),

$$(2\epsilon)^2 \text{cov}(W(s), W(t)) = \alpha(4\epsilon - 2\theta) + \alpha^2(-2e^{-\theta/\alpha} + e^{\frac{\theta-2\epsilon}{\alpha}} + e^{\frac{-\theta-2\epsilon}{\alpha}}),\ \theta \leq 2\epsilon.$$

Using series expansion of the Exponential functions, for small ϵ and $\theta \leq 2\epsilon$,

$$\begin{aligned}
&\alpha(4\epsilon - 2\theta) + \alpha^2(-2e^{-\theta/\alpha} + e^{\frac{\theta-2\epsilon}{\alpha}} + e^{\frac{-\theta-2\epsilon}{\alpha}}) \\
&= \alpha(4\epsilon - 2\theta) + \alpha^2[-2\{1 - (\theta/\alpha) + (\theta/\alpha)^2/(2!) - (\theta/\alpha)^3/(3!) + (\theta/\alpha)^4/(4!) - \cdots\} \\
&\quad + e^{-2\epsilon/\alpha}\{2 + 2(\theta/\alpha)^2/(2!) + 2(\theta/\alpha)^4/(4!) + \cdots\}] \\
&= (4\alpha\epsilon - 2\alpha^2 + 2\alpha^2 e^{-2\epsilon/\alpha}) + \alpha^2(e^{-2\epsilon/\alpha} - 1)(\theta/\alpha)^2 + (2\alpha^2/3!)(\theta\alpha)^3 \\
&\quad + (2\alpha^2/4!)(e^{-2\epsilon/\alpha} - 1)(\theta/\alpha)^4 + (2/5!)(\theta/\alpha)^5 + \cdots \\
&= a_0 + a_1\theta^2 + (\theta^3/3\alpha) + O(|\theta|^4) \text{ in a neighborhood of } 0.
\end{aligned}$$

Thus,

$$\text{cov}(W(s), W(t)) = b_0 + b_1\theta^2 + b_2\theta^3 + O(\theta^4) \text{ as } \theta \downarrow 0 \tag{B.1}$$

for appropriate real constants b_0, b_1, and b_2 depending on α and ϵ. Theorem 2 of Stein (1999) and (B.1) imply that the covariance function of W has 2 derivatives at the origin.

APPENDIX C

SUPPLEMENT TO THE DERIVATION OF EQUATION (2.7) IN SECTION 2

Recall that the mean square smoothness of a process in the Matérn class through its behavior at the origin. Consider $Z(t)$ and $Z(s)$ on $\mathcal{S}^1$.

$$K(\theta;\sigma^2,\alpha,\nu)=\frac{\pi^{1/2}\sigma^2}{2^{\nu-1}\Gamma(\nu+1/2)}(\theta/\alpha)^{\nu}K_{\nu}(\theta/\alpha),\quad \theta\geq 0, \tag{C.1}$$

$$K(\theta;\sigma^2,\alpha,0.5)=\pi\sigma^2\alpha e^{-\theta/\alpha}. \tag{C.2}$$

For $0<\nu\leq 0.5$,

$$\mathrm{cov}(W(s),W(t))=a_0+a_1\theta^2+c_0\theta^{2(\nu+1)}+O(\theta^4)\text{ as }\theta\downarrow 0 \tag{C.3}$$

for $c_0=\frac{-\pi\sigma^2}{\Gamma(2(\nu+1)+1)\sin((\nu+1)\pi)2\epsilon^2}$ and appropriate real constants a_0 and a_1 depending on σ^2, ν, α, and ϵ. Theorem 2 of Stein (1999) and (C.3) imply that the covariance function of W has 2 derivatives at the origin.

For K as in (C.1) with ν not an integer and $m<\nu<m+1$,

$$K(\theta)=\sum_{j=0}^{m}b_j\theta^{2j}-\frac{\pi\sigma^2}{\Gamma(2\nu+1)\sin(\nu\pi)}\theta^{2\nu}+O(\theta^{2m+2})\text{ as }\theta\downarrow 0 \tag{C.4}$$

for appropriate real constants $b_0,\ldots,b_m$ depending on σ^2, ν, and α.

Consider

$$\int_{t-\epsilon}^{t+\epsilon}\int_{s-\epsilon}^{s+\epsilon} K(|u-v|)dudv = \underbrace{\int_{t-\epsilon}^{t+\epsilon}\int_{s-\epsilon}^{s+\epsilon}\sum_{j=0}^{m} b_j|u-v|^{2j}dudv}_{\text{(i)}}$$
$$\underbrace{-\int_{t-\epsilon}^{t+\epsilon}\int_{s-\epsilon}^{s+\epsilon}\frac{\sigma^2\pi}{\Gamma(2\nu+1)\sin(\nu\pi)}|u-v|^{2\nu}dudv}_{\text{(ii)}}.$$

For relatively small ϵ and $\theta = |t-s| \le \epsilon$ $(s<t)$,

$$\int_{t-\epsilon}^{t+\epsilon}\int_{s-\epsilon}^{s+\epsilon}|u-v|^{2j}dudv$$

$$\begin{aligned}
&= \int_{t-\epsilon}^{s+\epsilon}\Big[\int_{s-\epsilon}^{v}(v-u)^{2j}du + \int_{v}^{s+\epsilon}(u-v)^{2j}du\Big]dv + \int_{s+\epsilon}^{t+\epsilon}\int_{s-\epsilon}^{s+\epsilon}(v-u)^{2j}dudv \\
&= \int_{t-\epsilon}^{s+\epsilon}\frac{1}{2j+1}\{(v-s+\epsilon)^{2j+1} + (s+\epsilon-v)^{2j+1}\}dv \\
&\quad + \int_{s+\epsilon}^{t+\epsilon}\frac{1}{2j+1}\{(v-s+\epsilon)^{2j+1} - (v-s-\epsilon)^{2j+1}\} \\
&= \frac{1}{(2j+1)(2j+2)}\{(2\epsilon+\theta)^{2j+2} + (2\epsilon-\theta)^{2j+2} - 2\theta^{2j+2} - 2(2\epsilon)^{2j+2}\}.
\end{aligned}$$

$$\begin{aligned}
\text{(i)} &= \sum_{j=0}^{m} b_j \frac{1}{(2j+1)(2j+2)}\underbrace{\{(2\epsilon+\theta)^{2j+2} + (2\epsilon-\theta)^{2j+2} - 2\theta^{2j+2} - 2(2\epsilon)^{2j+2}\}}_{(*)} \\
&= \sum_{j=0}^{m} b_j'\theta^{2j} + b_j''\theta^{2j+2} + O(\theta^{2(m+1)+2}) = \sum_{j=0}^{m+1} a_j\theta^{2j} + O(\theta^{2(m+1)+2})
\end{aligned}$$

for appropriate real constants $a_0, \ldots, a_{m+1}$ depending on σ^2, ν, α, and ϵ,

where

$$
\begin{aligned}
(*) =&(2\epsilon)^{2j+2} + (2j+2)(2\epsilon)^{2j+1}\theta + (2j+2)(2j+1)(2\epsilon)^{2j}\theta^2/(2!) + \cdots \\
&+ (2\epsilon)^{2j+2} - (2j+2)(2\epsilon)^{2j+1}\theta + (2j+2)(2j+1)(2\epsilon)^{2j}\theta^2/(2!) - \cdots \\
&- 2\theta^{2j+2} - 2(2\epsilon)^{2j+2} \\
=&2(2j+2)(2j+1)(2\epsilon)^{2j}\theta^2/(2!) \\
&+ 2(2j+2)(2j+1)(2j)(2j-1)(2\epsilon)^{2j-2}\theta^4/(4!) + \cdots,
\end{aligned}
$$

and

$$
\begin{aligned}
\text{(ii)} =& \underbrace{\frac{\pi\sigma^2}{\Gamma(2\nu+1)\sin(\nu\pi)}\frac{1}{(2\nu+1)(2\nu+2)}}_{=c} \\
&\times \{(2\epsilon+\theta)^{2\nu+2} + (2\epsilon-\theta)^{2\nu+2} - 2\theta^{2\nu+2} - 2(2\epsilon)^{2\nu+2}\} \\
=&c \times 2(2\nu+2)(2\nu+1)(2\epsilon)^{2\nu}\frac{\theta^2}{2!} + 2(2\nu+2)(2j+1)(2\nu)(2\nu-1)(2\epsilon)^{2\nu-2}\frac{\theta^4}{4!} \\
&+ \cdots - c \times 2\theta^{2\nu+2} \\
=&\sum_{j=0}^{m+1} c_j\theta^{2j} - 2c\theta^{2(\nu+1)} + O(\theta^{2(m+1)+2}) \text{ in a neighborhood of } \theta = 0
\end{aligned}
$$

for appropriate real constants $c_0, \ldots, c_{m+1}$ depending on σ^2, ν, and ϵ.

Thus, for $m < \nu < m+1$,

$$
\int_{t-\epsilon}^{t+\epsilon}\int_{s-\epsilon}^{s+\epsilon} K(|u-v|)dudv = \sum_{j=0}^{m+1} a_j\theta^{2j} + c_0\theta^{2(\nu+1)} + O(\theta^{2(m+1)+2}) \text{ as } \theta \downarrow 0 \quad \text{(C.5)}
$$

for $c_0 = \frac{-\pi\sigma^2}{\Gamma(2\nu+1)\sin(\nu\pi)} \times \frac{-2}{(2\nu+1)(2\nu+2)} = \frac{-2\pi\sigma^2}{\Gamma(2(\nu+1)+1)\sin((\nu+1)\pi)}$ and appropriate real constants $a_0, \ldots, a_{m+1}$ depending on σ^2, ν, α, and ϵ. Theorem 2 of Stein (1999) together

with (C.5) implies that $\int_{t-\epsilon}^{t+\epsilon}\int_{s-\epsilon}^{s+\epsilon} K(|u-v|)dudv$ is $2(m+1)$ times differentiable if and only if $\nu > m$. Thus, $\nu \in (0, 0.5]$, the covariance function of W has 2 derivatives.